

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>7</sup> : <b>B01J 19/00</b>	<b>A2</b>	(11) International Publication Number: <b>WO 00/67894</b> (43) International Publication Date: 16 November 2000 (16.11.00)
<p>(21) International Application Number: <b>PCT/US00/12825</b></p> <p>(22) International Filing Date: <b>10 May 2000 (10.05.00)</b></p> <p>(30) Priority Data: <b>60/134,028</b>      <b>12 May 1999 (12.05.99)</b>      <b>US</b></p> <p>(71) Applicant (for all designated States except US): <b>AXYS PHARMACEUTICALS, INC. [US/US]; 180 Kimball Way, South San Francisco, CA 94080 (US).</b></p> <p>(72) Inventors; and (75) Inventors/Applicants (for US only): <b>STEWART, Michael [GB/US]; 12742 Torrey Bluff #36, San Diego, CA 92130 (US). NANTHAKUMAR, Alaganadan [US/US]; 6737 Withesail Street, Carlsbad, CA 92009 (US). WATSON, Andrew [GB/US]; 531 Colorado Avenue, Palo Alto, CA 94305 (US).</b></p> <p>(74) Agent: <b>SHERWOOD, Pamela, J.; Bozicevic, Field &amp; Francis LLP, Suite 200, 200 Middlefield Road, Menlo Park, CA 94025 (US).</b></p>	<p>(81) Designated States: <b>AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.</b> ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p><b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i></p>	
(54) Title: <b>METHODS OF SOFTWARE DRIVEN FLOW SORTING FOR REITERATIVE SYNTHESIS CYCLES</b>		
(57) Abstract		
<p>Methods are provided for synthesizing libraries of complex organic molecules on labeled particles. A set of particles encoded with varying levels and combinations of dyes, which provide a detectable address, are used as the support for organic synthesis. The addresses of a set of particles is read by flow cytometry, and used to classify the microspheres. The set of microspheres is then sorted into groups by flow cytometry, using a modified look up table. Monomers are coupled to each microsphere in a group, where each group corresponds to a different coupling reaction. The groups are then combined and resorted, and a second round of addition reactions performed. The reiterative process of sorting into groups; and coupling additional monomers to the growing oligomer chain is performed for sufficient rounds to provide an oligomer of the desired length. The resulting "liquid array" is a set of encoded microspheres comprising a library of synthesized oligomers, where each sequence in the oligomer library corresponds to a distinct address of fluorescent output data.</p>	<p>— Fluorochrome 1      — Fluorochrome 2</p>	

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Larvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## METHODS OF SOFTWARE DRIVEN FLOW SORTING FOR REITERATIVE SYNTHESIS CYCLES

### INTRODUCTION

#### 5 Background

Solid phase synthesis of complex organic molecules such as polypeptides and nucleic acids is the overwhelming method of choice for producing many compounds used in research.

The availability of these synthetic reactions has permitted a wide variety of compounds and variations to be produced. In recent years the synthesis and use of "arrays" built on solid phase  
10 substrates has exploded. The term "array" in this context is used to indicate a set of target compounds having distinct sequences, where each target compound is coded for identification. One example of coding is the use of "tags", where the target compounds are attached to detectable label, or tag. The tag provides coded information about the sequence.

Another example is spatial coding, where the position of the molecule is fixed, and that  
15 position is correlated with the sequence. One form of these arrays is produced by reiterative synthesis cycles, where a compound such as a polypeptide or oligonucleotide is synthesized *in situ* at precise locations on a planar substrate. Typically, subunits, or monomers, are added sequentially to the target compound in rounds of addition reactions, where different sequences of the target compound are achieved by varying the order in which monomers are applied.

For example, photolithography has been combined with solid phase DNA synthesis for  
20 the construction of high-density DNA probe arrays. Synthetic linkers modified with photochemically removable protecting groups are attached to a glass substrate, and light through a photolithographic mask is applied to produce localized photodeprotection. Deoxynucleosides are coupled to the deprotected sites. In reiterative rounds, different regions are deprotected for  
25 coupling. By using different masking strategies, specified sequences are produced at particular locations in the array.

Oligonucleotide arrays built on solid supports are being used for the analysis of sequence variation for scoring and identification of polymorphisms, and for expression profiling. For example, arrays have been developed for analysis of HIV sequence variation, genotyping of  
30 cytochrome P450 variants and BRCA1 re-sequencing. Advantages of oligonucleotide arrays over conventional approaches are numerous: data capture is automatic and can be interpreted using simple heuristics, samples are easy to process, multiplexing of many data points per sample is possible, and the arrays are created using DNA sequence information rather than any need for physical clones. Current embodiments of microarrays do not, however, allow many samples to  
35 be processed in parallel.

A major use for arrays is in DNA analysis and diagnostic testing. These panels are likely to comprise many hundreds or thousands of data points. For example, the number of mutations

found in the cystic fibrosis transmembrane receptor gene is greater than 200, and CF is a monogenic disease. Detection of the appropriate sequence variants or gene expression measurements is likely to require the analysis of many hundreds of thousands of loci and gene expression levels in many thousands of patients. Discovery of these diagnostic and prognostic markers is currently a major bottleneck in making DNA diagnostic testing more widespread. The ideal technology platform for these analyses, then, would be one that used genetic information to develop DNA diagnostic panels, which will be used as the method of choice in a high-throughput diagnostic setting.

#### 10 *Relevant Literature*

Methods for the manufacture and use of spatial arrays of oligonucleotides and polypeptides are widely known. Recent reviews include Lipshutz *et al.* (1999) Nat. Gen. suppl. 21:20-24; Debouck and Goodfellow (1999) Nat. Gen. suppl. 21:48-50; and Hacia (1999) Nature. Gen. Suppl. 21:42-47. Representative patents include US 5,700,637, issued Dec. 23, 1997; US 5,744,305, issued April 28, 1998; and US 5,800,992, issued Sept. 1, 1998.

The use of a multiplexed microsphere set for analysis of clinical samples by flow cytometry is described in International Patent application no. 97/14028; and Fulton *et al.* (1997) Clinical Chemistry 43:1749-1756). The specific use of this method for analysis of DNA samples may be found in U.S. Patent no. 5,736,330, issued April 7, 1998. Hakala *et al.* (1997) Bioconjugate Chem 8:378-384 describe oligonucleotide hybridization detection on microspheres. DNA fragment sizing and sorting by laser-induced fluorescence is described in U.S. Patent no. 5,558,998, issued September 24, 1996.

The materials and techniques used in combinatorial chemical techniques are known in the art, see for example Houghten (1985) Proc. Natl. Acad. Sci. USA 82:5131-5135; Geysen *et al.* (1984) Proc. Natl. Acad. Sci. USA 81:3998-4002; Pirrung *et al.* (1995), J. Am. Chem. Soc. 117:1240-1245; Smith *et al.* (1994) BioMed Chem. Lett. 4:2821-2824; Beaucage *et al.* (1981) Tetrahedron Lett. 22:1859-62, and Itakura *et al.* (1975) J. Biol. Chem. 250:4592 (1975).

Methods of defining data point clusters in N-dimensional space are described by Bierre *et al.*, U.S. Patent no. 5,627,040; and U.S. Patent no. 5,739,000. Other algorithms for clustering cells in flow cytometry are described in Verwer *et al.*, U.S. Patent no. 5,605,805. Keij *et al.* (1995) Cytometry 19(1):92-6, 1995 uses look up tables to perform a classification of particles in a flow cytometer. Data handling in flow cytometry is discussed in Frankel *et al.* (1996) Cytometry 23(4):290-302; van den Engh and Stokdijk (1989) Cytometry 10(2):282-93; Murphy (1985) Cytometry 6(4):302-9; Bakker Schut *et al.* (1993) Cytometry 14(6):649-59; and Zilmer *et al.* (1995) Cytometry 20(2):102-17.

Methods for cluster analysis, principle components analysis and optimization are described in Massart *et al.* (1998), "Data Handling in Science and Technology – Volume 2. Chemometrics: a textbook", Elsevier. Computer algorithms for principle components analysis (or factor analysis) and optimization may be found in Press *et al.* (1989) "Numerical Recipes in  
5 Pascal", Cambridge University Press.

#### SUMMARY OF THE INVENTION

Methods are provided for synthesis of a library of complex organic molecules on labeled particles, using a reiterative split and pool approach. Initially, a set of particles are encoded with  
10 varying levels and combinations of labels, e.g. fluorescent dyes, which provide an identifier, or address, for each microsphere. The address profile for a set of encoded particles is read by flow cytometry. The address information is analyzed by a combination of algorithms, and used to classify the microspheres. Look up tables are generated from this data. The set of microspheres is then sorted into groups by flow cytometry, using the look up tables and fluorescence output  
15 data.

Monomers are then coupled to each microsphere in a group, using the appropriate chemistry for the oligomer to be synthesized. Each group may have a different monomer. The groups are then combined and resorted, and a second round of addition reactions performed. The reiterative process of sorting into groups; and coupling additional monomers to the growing  
20 oligomer chain is performed for sufficient rounds to provide an oligomer of the desired length. The subject methods find particular use in the synthesis of highly complex sets of oligomers comprising greater than  $10^3$  different oligomer sequences. The set of oligomers bound to encoded microspheres are useful in a variety of methods, particularly binding or affinity studies.

#### 25 BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a graph depicting spectral overlap for two fluorescent dye excited by a single laser.

Figure 2 is a plot of a two fluorescent dye microsphere set containing 64 distinct microsphere addresses.

30 Figure 3 is a flow diagram of histogram analysis of classification channels with spectral overlap present.

Figures 4A-4C illustrate a histogram analysis of channel C1 of data shown in Figure 2. Figure 4A shows the start of the first bin. Figure 4B shows end of first bin. Figure 4C. Shows the limits of all bins.

35 Figure 5 illustrates a histogram of C2 for data shown in Figure 2 after segmenting microspheres according to bins shown in Figure 4C. The histogram is for the eighth bin.

Figures 6A and 6B illustrate a four processor parallel DSP system to classify the 64 microsphere set. Figure 6A shows the assignment of microsphere clusters to processors. Figure 6B is a diagram showing 4 DSP units with dual ported shared memory (M1), local memory (M2) accessed by the DSP via a local bus. Communication with the host is via high-speed serial communications lines to transfer microsphere cluster assignments and the shared memory bus to transfer the microsphere data from the instrument interface to all DSP processors simultaneously.

Figures 7A and 7B are an example of simple lookup table (LUT). Figure 7A shows the lookup table programmed with desired destination of microspheres with a measured value equal to the address shown. Figure 7B shows the LUT implemented using a small memory device.

Figures 8A to 8C show a simple bitmap using 2 channels, each digitized to 4 levels. Figure 8A is an assignment of sorting destination by C1 and C2 values. Figure 8B is a conversion of bitmap to linear LUT. Figure 8C is an implementation of bitmap using a small memory device.

Figures 9A to 9C are an example of hierarchical LUTs for four channels. Figure 9A shows bitmaps containing cluster IDs for clustering by C1/C2 and C3/C4. Figure 9B is an assignment of first base of sequence to clusters. Figure 9C illustrates implementation of hierarchical LUT using memory devices.

Figure 10 is a table illustrating the assignment of cluster ID numbers with nucleotide sequences.

Figure 11 shows a synthetic scheme to separate dye binding and oligomer synthesis.

Figure 12 illustrates the coupling of linkers to surface amines on microspheres.

Figure 13 is a capillary electropherogram of 20-mer sequence synthesized on polystyrene.

Figure 14 is a histogram depicting the effect of oligomer synthesis conditions on a mixture of microspheres.

Figure 15 is a histogram analysis of sorting results from mixtures of microspheres.

Figure 16 shows 2 dimensional dot-plots of a mixture of microspheres comprising oligomer sequences.

Figure 17 is a graph depicting hybridization of microspheres comprising oligomers to detect single nucleotide polymorphisms.

Figure 18 is a bar graph representing the results of different hybridization conditions.

#### DESCRIPTION OF THE SPECIFIC EMBODIMENTS

Methods are provided for synthesis of a library of complex organic molecules on labeled particles, using a reiterative split and pool approach. Initially, a set of particles are encoded with

varying levels and combinations of labels, e.g. fluorescent dyes, which provide an identifier, or address, for each microsphere. The address profile for a set of encoded particles is read by flow cytometry. The address information is analyzed by a combination of algorithms, and used to classify the microspheres. Look up tables are generated from this data.

- 5       The set of microspheres is then sorted into groups by flow cytometry, using the look up tables and fluorescence output data generated during the sort. Monomers are then coupled to each microsphere in a group, using the appropriate chemistry for the oligomer to be synthesized. Each group may be coupled to a different monomer. The groups are then combined and resorted, and a second round of addition reactions performed. The reiterative process of sorting  
10       into groups; and coupling additional monomers to the growing oligomer chain is performed for sufficient rounds to provide an oligomer of the desired length.

- The resulting "liquid array" is a set of encoded microspheres comprising a library of synthesized oligomers. Each sequence in the oligomer library corresponds to a distinct address of fluorescent output data. The identity of the oligomer sequence is read out by flow cytometry,  
15       where the microsphere provides an identifier for the oligomer. The combined microsphere and oligomer may be used in assays, or the oligomer may be cleaved from the microsphere and used separately.

- The primary sorting means is by flow cytometry. The flow cytometer analyzes individual microspheres by size and fluorescence, distinguishing fluorescent colors, which may include  
20       green (530 nm), orange (585 nm) and red (>650 nm), simultaneously. Microsphere size, determined by 90-degree light scatter, is used to eliminate microsphere aggregates from the analysis. Internal ratios of fluorescence are used for microsphere classification, and additional colors may be used for analyte measurement. Currently available instruments are very rapid, with detection rates up to  $10^5$  particles per second.

- 25       The software required for the present methods consists of two components. The first module is used for classification of the "empty" microsphere set, i.e. microspheres that do not have attached oligomers. The microsphere set is analyzed by flow cytometry, and an output file generated that is based on the measured intensities of the encoding fluorescent dyes. This information is combined with the oligomer synthesis parameters that define the specific sequence  
30       combinations to generate a series of rules for sorting during the reiterative synthesis steps. A similar module is used to analyze assay data. The parameters used to sort microspheres may also be used to demultiplex assay data.

- The liquid array find particular use in binding or affinity assays. In such assays, the liquid array is contacted with a sample suspected of containing a binding member for one or more of  
35       the oligomers in the array. For example, nucleic acid samples may be contacted with a DNA array, or with a peptide array to detect specific binding. Preferably, the sample comprises a detectable label distinct from the liquid array encoding labels. The unbound sample is washed



away from the array. The liquid array is then analyzed by flow cytometry. The presence of sample material is detected, and the bound oligomer-microsphere is identified by reading the encoded fluorescent address. The use of flow cytometry is well suited for complex arrays, e.g. those comprising more than  $10^3$  different oligomer sequences. The methods are also well suited for analyzing large numbers of samples, as a flow cytometer can readily analyze several hundred samples or more in an hour.

In one embodiment of the invention, a means is provided for analyzing nucleic acids by hybridization to oligonucleotides. These oligonucleotides are synthesized on microspheres having an encoded address that can be uniquely identified in a flow cytometer. Following synthesis, a suitably labeled test nucleic acid is hybridized to the mixture of microspheres. The mixture is then read in the flow cytometer where the class of each microsphere is identified, and therefore the oligonucleotide sequence that is attached; as well as the degree of hybridization of the test nucleic acid. By choosing appropriate oligonucleotide sequences to be synthesized, and by appropriate labeling of the test nucleic acid, single base pair genotyping, mRNA quantitation, and analysis of sequence variation can be detected.

#### Definitions

It is to be understood that this invention is not limited to the particular methodology, protocols, cell lines, animal species or genera, and reagents described, as such may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

As used herein the singular forms "a", "and", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a cell" includes a plurality of such cells and reference to "the protein" includes reference to one or more proteins and equivalents thereof known to those skilled in the art, and so forth. All technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs unless clearly indicated otherwise.

**Liquid array:** as used herein refers to a library of oligomers having defined sequences, where the oligomers are attached to microspheres or microspheres having an encoded, detectable address. Each oligomer sequence is associated with a distinct address. The address is read at the single particle level, enabling identification of the associated oligomer sequence.

Oligomers made according to the present invention are synthesized on the microspheres by repeated cycles of addition reactions. The oligomers may be bound to the microspheres through a covalent or a high affinity non-covalent linkage, usually a covalent linkage. Nucleic

acid oligomers are of particular interest, including DNA, RNA, PNA, and analogs thereof. Polypeptide oligomers are also synthesized by the provided methods.

The encoded address is generally provided by a combination of fluorescent dyes, where the presence of a specific combination at pre-determined intensities provides a unique signature output when excited with light of a particular wavelength. Conveniently, the address is read with a flow cytometer, which provides for both analysis and sorting functions. The use of lasers in particle sorting is well known in the art, and suitable machines having from one to three lasers are commercially available (Becton Dickinson). A collection of measured fluorescent intensities from a liquid array will be referred to as a data array.

The "complexity" of the liquid array is intended to refer to the number of distinctly addressed oligomer-microsphere address combinations. For some purposes, one address will correspond to one oligomer sequence. However, the invention is not limited to this relationship. Multiple, particularly degenerate, oligomer sequences may be used with a single address. Conversely, an "address" may be broadly defined, such that one oligomer sequence is associated with a group of addresses, rather than a single point.

Arrays of particular interest have a complexity of at least about  $10^2$  distinct oligomer-microsphere address combinations, and are usually at least about  $10^3$  combinations. In some embodiments, the complexity will be at least about  $10^4$  combinations, and may be as much as  $10^5$  or  $10^6$  combinations. An important factor in complexity is the ability to accurately and reproducibly discriminate individual addresses in a mixture. The present methods provide algorithms for such analysis.

*Microspheres:* microspheres provide an insoluble support, and encoded address, for synthesis and detection of oligomers. A wide variety of polymers can be employed in microsphere fabrication, including latex, polystyrene, polyethylene/polypropylene, polycarbonate, polymethylmethacrylate, chloromethylpolystyrene-1%-divinylbenzene, silica, porous glass, aluminosilicates, borosilicates, metal oxides such as alumina and nickel oxide, and various clays.

Polystyrene is a preferred polymer. Included in the term "polystyrene" are polymers that have been substituted to some extent with substituents that are not capable of reaction under the conditions used for synthesis, including, for example, alkyl substituents such as methyl, ethyl, propyl, butyl, alkoxy substituents, etc. In order to increase the stability and insolubility in organic solvents, polystyrene resins that have been cross-linked by co-polymerization with at most 5 mol%, and preferably from about 1 to 2 mol% with divinyl benzene or butadiene are also used.

The polymer should be substantially insoluble in the reaction solvents employed and relatively chemically inert to the reagents employed during processing, except for the chemical reactivity required to form a chemical bond with the initial monomer through which the oligomer

is attached to the support. Suitable microspheres are tolerant of the solvents used in organic synthesis of the oligomers, and are compatible with the fluorescent dyes used for encoding.

The linkage between the microsphere and oligomer may be any suitable functionality appropriate for the oligomer synthetic chemistry. A large number of heterofunctional compounds are available for linking to entities. Illustrative entities include: azidobenzoyl hydrazide, N-[4-(p-azidosalicylamino)butyl]-3'-[2'-pyridyldithio]propionamide), bis-sulfosuccinimidyl suberate, dimethyladipimide, disuccinimidyltartrate, N-γ-maleimidobutyryloxysuccinimide ester, N-hydroxy sulfosuccinimidyl-4-azidobenzoate, N-succinimidyl [4-azidophenyl]-1,3'-dithiopropionate, N-succinimidyl [4-iodoacetyl]aminobenzoate, glutaraldehyde, NHS-PEG-MAL, Shearwater polymers, and succinimidyl 4-[N-maleimidomethyl]cyclohexane-1-carboxylate; 3-(2-pyridyldithio)propionic acid N-hydroxysuccinimide ester (SPDP) or 4-(N-maleimidomethyl)-cyclohexane-1-carboxylic acid N-hydroxysuccinimide ester (SMCC).

The microspheres are encoded by attachment or encapsulation of a fluorescent agent, quantum dots or heavy metal complexes. Suitable fluorescent dyes for encoding are known in the art, including fluorescein isothiocyanate (FITC), rhodamine and rhodamine derivatives, Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein (JOE), 6-carboxy-X-rhodamine (ROX), 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA), sulfonated rhodamine, etc.

The number of possible addresses in a microsphere set depends on the number of dyes used, and the number of fluorescence levels that can be discriminated. With 2 dyes and 8 possible fluorescence levels for each,  $8 \times 8 = 64$  different microsphere sets can be discriminated. With 6 dyes, and 10 levels,  $10 \times 10 \times 10 \times 10 \times 10 \times 10 = 1 \times 10^6$  addresses can be discriminated.

*Microsphere set:* is used to refer to the microsphere component of a liquid array or to the set of microspheres prior to performance of the synthetic reactions. The microspheres are encoded with the desired range of coding labels. A set of microspheres will typically include a plurality of different addresses. The data output from a microsphere set may be referred to as a data set.

*Microsphere cluster:* When the output data is read from a group of microspheres having a particular address, the data points for microspheres with identical encoding will form a cluster around the average fluorescent intensities for that type of microsphere, herein referred to as a microsphere cluster. Microspheres with different addresses should belong to different clusters.

The set of data points for a microsphere cluster may be referred to as a data cluster.

*Flow Cytometry:* Flow cytometry methodologies are well developed for cell-based assays and cell sorting. Clinical flow cytometers are common instruments worldwide. High speed sorting flow cytometers are presently made with the ability to measure 7 different fluorescence parameters, from 3 different laser excitations, from a particle as it flows past the flow cell. These instruments operate by flowing cells past a detector and taking measurements, typically fluorescence and size, on each cell or particle. The particles can also be sorted based on the fluorescence and size measurements read from the particle.

If the signals are obtained with multiple excitation beams, the pulses from a single particle will reach different detectors at different times. The asynchronous events can be correlated either before or after the pulse digitization. One approach to pre-processing synchronization is to hold the pulse values in analog circuits until all measurements of an event have been completed. After the event leaves the last measurement beam, the held values are input to AD converters. A more efficient approach is to delay the earliest pulses with analog delay lines such that all signals enter the acquisition channels simultaneously. The cycle time is the AD conversion time plus the pulse width.

U.S. Patent no. 5,150,313, van den Engh, *et al.* describe a digitally synchronized, parallel pulse processing and data acquisition system. Parallel pulse processing is achieved by equipping each input channel with a set of pulse processing electronics. The detector pulses are immediately converted into digital values which are temporarily stored in first in, first out (FIFO) buffers which are connected to a digital data bus. Digital timing circuitry keeps track of the stored values. After a particle has traversed all illumination beams its measured values are transferred as a package to the acquisition computer over the data bus. The cycle time is determined by the length of the AD conversion process alone. Since the channel has processed the input signals independently, the scheme is extended to any number of input channels and illumination beams.

The term *oligomer* is used herein to indicate a chemical entity that contains a plurality of monomers. The terms "oligomer" and "polymer" may be used interchangeably. Examples of oligomers and polymers include polypeptides, polydeoxyribonucleotides, polyribonucleotides, protein nucleic acids, other polynucleotides which are N- or C-glycosides of a purine or pyrimidine base, polysaccharides, and other chemical entities that contain repeating units of like chemical structure.

*Monomer* as used herein refers to a chemical entity that can be covalently linked to one or more other such entities to form an oligomer. Examples of "monomers" include amino acids, nucleotides, saccharides, peptoids, and the like. In general, the monomers used in conjunction with the present invention have first and second sites, e.g. C-termini and N-termini, or 5' and 3' sites, suitable for binding to other like monomers by means of standard chemical reactions, e.g.

condensation, nucleophilic displacement of a leaving group, or the like, and typically, a diverse element that distinguishes a particular monomer from a different monomer of the same type, e.g. an amino acid side chain, a nucleotide base, etc. An initial support-bound monomer is used as a building-block in a multi-step synthesis procedure to form an oligomer, such as in the synthesis of oligopeptides, oligopeptoids, oligonucleotides, and the like. In some cases, the building block for oligomers will be a dimer, trimer, or other multimeric form.

*Nucleic Acid Synthesis:* Methods for preparing oligonucleotides are known in the art. For example, oligonucleotides can be prepared using conventional phosphoramidite chemistry. In a typical phosphoramidite synthesis, a reactive 3' phosphorous group of one nucleoside is coupled to the 5' hydroxyl of another nucleoside. The former is a monomer, delivered in solution as a 5' hydroxyl protected phosphoramidite derivative; the latter is immobilized on a solid support as a 5' hydroxyl protected derivative. The first step of the synthesis cycle is deprotection, in which the protecting group on the immobilized nucleoside is removed to free the 5' hydroxyl group for the coupling reaction. The next step is coupling, in which an activated intermediate is created by simultaneously adding the protected phosphoramidite derivative and a weak acid, e.g., tetrazole. The acid protonates the nitrogen of the phosphoramidite derivative, making it susceptible to nucleophilic attack. Finally, the internucleotide phosphite linkage is converted to the more stable phosphotriester linkage by oxidizing, e.g., with iodine solution. After the oxidation, the 5' hydroxyl protecting group is removed and the cycle is repeated until chain elongation is complete. Suitable 5' hydroxyl protecting groups include dimethoxytrityl. Deprotection is effected by any means which removes the protecting group and gives the desired product in reasonable yield. For example, detritylation can be effected with trifluoroacetic acid.

Another nucleic acid of interest is peptide nucleic acid (PNA), (see U.S. Patent no. 5,539,082, Nielsen, et al.) Peptide nucleic acids are synthesized by adaptation of standard solid phase peptide synthesis procedures. The monomers are amino acids or their activated derivatives, protected by standard protecting groups. The oligonucleotide analogs also can be synthesized by using the corresponding diacids and diamines.

The PNA monomers are protected at the reactive primary amino groups and the exocyclic amines of the bases. The carboxyl group of each monomer is unprotected and is activated prior to coupling. There are two different chemical methods for synthesis of PNA: Fmoc (9-fluorenylmethoxycarbonyl) method; and the tBoc (tert-butyloxycarbonyl) method.

The Fmoc protection of the support or support-bound monomer is removed with a basic (piperidine) solution to free the amino group for coupling. The carboxyl group of the next monomer is activated using a mixture of HATU, DIPEA, and lutidine. The activated monomer is coupled to the growing chain by formation of an amide bond. Excess reagents and high concentrations are used to drive reactions as close to completion as possible. Unreacted chains

(failure sequences) are capped with an acetylating solution to prevent further elongation. These four steps are repeated until the PNA oligomer is fully assembled. If desired, the PNA is cleaved from the support and the base protecting groups are removed by treatment with a mixture of TFA and m-cresol. PNA may be synthesized on a polyethylene glycol-polystyrene (PEG-PS) support with a PAL linker. The PAL linker yields a PNA amide upon cleavage of the final product.

Synthesis by the tBoc method employs different reagents. The tBoc group protects the primary amino groups of the supports and monomers. The bases are protected with the benzyloxycarbonyl (Z) group. The support matrix is PEG-PS with a BHA linker that yields an amide upon cleavage.

10

The terms "*nucleoside*" and "*nucleotide*" are intended to include those moieties which contain not only the known purine and pyrimidine bases, but also other heterocyclic bases that have been modified. Such modifications include methylated purines or pyrimidines, peptide nucleic acids, acylated purines or pyrimidines, or other heterocycles. In addition, the terms "*nucleoside*" and "*nucleotide*" include those moieties which contain not only conventional ribose and deoxyribose sugars, but also other sugars as well. Modified nucleosides or nucleotides will also include modifications on the sugar moiety, e.g. wherein one or more of the hydroxyl groups are replaced with halogen, aliphatic groups, or are functionalized as ethers, amines, or the like.

20

*Peptide synthesis:* Solid-phase peptide synthesis involves the successive addition of amino acids to create a linear peptide chain. The C-terminus of the growing peptide is covalently bound to the microsphere during synthesis. Three chemical reactions are repeated for each amino acid that is added to the peptide chain: deprotection, activation, and coupling.

Protected amino acids are derivatized to prevent unwanted reactions at their alpha-amino and side-chain functionalities. During deprotection, the protecting group is removed to make the alpha-amino group on the end of the peptide chain accessible. Activation converts the next amino acid to be added to an active ester. During coupling, the active ester forms an amide bond with the deprotected alpha-amino group on the end of the peptide chain. After coupling, a new cycle of synthesis begins with the next deprotection. When synthesis is complete, chemical cleavage removes side-chain protecting groups from the peptide (Merrifield (1963) J. Am. Chem. Soc. 85:2149-2154).

The terms "*protection*" and "*deprotection*" as used herein relate, respectively, to the addition and removal of chemical protecting groups using conventional materials and techniques within the skill of the art and/or described in the pertinent literature. Protecting groups prevent the site to which they are attached from participating in the chemical reaction to be carried out. Methods and conditions for the removal of protecting groups are well known in the art.

Any number of protecting groups can be used, as will be appreciated by those skilled in the art. Suitable protecting groups will be known to or easily deduced by those working in the field of synthetic organic or bio-organic chemistry. The only requirements for the protecting groups used herein are that they be "orthogonal" so as to remain in place during other chemical  
5 syntheses or procedures which are carried out on the unprotected sites, e.g., coupling of amino acids, peptide mimetics, nucleotides, and the like; and they are compatible with whatever temperatures, reaction conditions and reagents are employed while they are in place, i.e., are not degraded, chemically altered, or removed from the protected site.

Frequently, although not necessarily, the protecting groups are acid-cleavable. Examples  
10 of suitable protecting groups include, but are not limited to: (a), for diol protection, 2,2-dimethoxypropane, acetals such as benzylidene acetal and *p*-methoxybenzylidene acetal, bifunctional silyl ethers such as di-*t*-butylsilylene, and compounds which upon reaction with a 1,2-diol will form acetonides, cyclic carbonates or cyclic boronates; and (b) for protection of a single hydroxyl site, (i) protecting groups which will give rise to ethers, e.g. tetrahydropyranyl,  
15 dihydropyranyl, trimethylsilyl, substituted or unsubstituted benzyl (if substituted, typically with electron withdrawing groups such as NO<sub>2</sub>), and triphenylmethyl, and (ii) protecting groups which will give rise to esters, such as acetyl, trifluoroacetyl, and trichloroacetyl.

#### CLASSIFICATION OF ENCODED MICROSPHERES

20 The first step in using a microsphere set for synthesis by sorting is to measure the fluorescence signals under standard operating conditions for the empty microsphere set. The fluorescence of dyes used to encode the microsphere addresses is measured in data channels labeled C1...CN. The flow cytometry instrument can measure a number of parameters for each microsphere. These include the forward and side light scatter (FSC and SSC) and one or more  
25 fluorescence wavelengths. The fluorescence measurements are traditionally labeled FL1, FL2 and so on.

As used in the methods of the present invention, from two to six of these fluorescent measurements (also referred to as fluorescent channels) will be used for encoding; and one or more fluorescent measurements will be used for determining the result of an assay. To  
30 distinguish the fluorescent measurements used to decode the microsphere identity from those used for determining the assay result, the encoding (or classification) channels are labeled as C1, C2 and so on. There will be one classification channel per fluorescent dye, which is labeled C1 to CN; where N is the number of fluorescent dyes used for classification. The channels used for determining assay results will be labeled FL1 and FL2. For example, if two fluorescent dyes  
35 are used to encode microspheres and two fluorescence probes are used for the assay, these would be measured using channels C1 and C2 for encoding and FL1 and FL2 for the assay.

The assignment of the channel labels to fluorescent dyes is arbitrary. To simplify the description of the software; the fluorescent dye used for encoding that emits at the shortest wavelength will be measured using the channel C1; while the fluorescent dye that emits at the next shortest wavelength will be measured using channel C2. For N fluorescent dyes the longest wavelength emission will be measured using channel CN.

A channel may be referred to as a channel number x, where x is the numerical part of the channel label. Similarly, statements about the relative order of two or more channels may be made based on the numerical part of the labels, for example a channel may have a smaller number than another. This will be used in explaining the histogram analysis method since the analysis will iterate over channels with increasing channel number.

The operator will optimize data channel gains to maximize the separation of the microsphere clusters. Once the gains have been set and recorded, a sample of microspheres from the microsphere set is measured and the data stored in an electronic format to be read by the classification software.

In the case where the same laser excites two dyes then it is expected that there will be spectral overlap from the lower channel to the higher channel. Though not relevant to the classification and sorting process forward and side scatter will be measured in data channels labeled FSC and SSC respectively. Correction of spectral overlap by subtracting a proportion of one signal from another will be referred to as compensation.

The classification process requires the software to identify the clusters present in a data set. The clusters can be characterized using the range of values for each channel CN, within which each microsphere belonging to the cluster is expected to lie. Alternatively, a characteristic point within a cluster can be defined and all microspheres closest to the characteristic point are regarded as belonging to that cluster. An automated process is used that will analyze multidimensional data with large numbers of clusters present.

Information is used to guide the clustering methods. The number of intensities levels for each dye used in address encoding is known, for example a 100 microsphere set may use 10 levels of two dyes. The clustering method of the present invention uses knowledge of the numbers of dye intensities to classify the microsphere set. This algorithm uses repeated application of histogram analysis.

The approach is based on the fact that lower channel numbers will show better separation than higher channels number due to spectral overlap broadening the peaks in the histogram. Therefore the histogram of C1 is analyzed first and the data points are binned by C1. The C2 histogram of each C1 bin is then analyzed. The data points are then binned by C2 to produce an initial classification. The k-means method is then applied to this initial classification to remove any anomalies that remained.



Spectral overlap occurs when two or more fluorescent dyes have emissions which are close together. This is illustrated in Figure 1, which shows the emission profiles of two fluorescent dyes as a function of wavelength. A single channel of the flow sorter measures the fluorescence that occurs between two wavelengths (to a first approximation). The wavelengths measured by two encoding channels are shown in the figure by the vertical lines. The wavelengths measured by channel C1 are between the first and second lines from the left. Similarly the wavelengths measured by channel C2 are between the third and fourth vertical lines. The area shaded in cross-hatched area represents the fluorescence from the first encoding fluorescent dye that is measured in channel C2 while the area shaded in gray represents the fluorescence from the second encoding fluorescent dye that is measured in channel C1.

The presence of fluorescence in a channel from another fluorescent dye (that is a fluorescent dye other than the intended one) is referred to as spectral overlap. This results in a systematic error in the fluorescence measured. The effect is generally more severe when the interfering fluorescence is from a fluorescent dye that emits at shorter wavelengths. In other words, the fluorescence from fluorescent dye 1 in channel C2 will be more than the fluorescence of fluorescent dye 2 in channel C1 for equal amounts of each fluorescent dye. By using the labeling convention mentioned above this means that there is expected to be more severe overlap from C1 into C2 and C2 into C3 and so on.

The amount of overlap decreases rapidly as the channels become separated in wavelength. That is, in this example, if there were a channel C3 at longer wavelengths than channel C2 the overlap from fluorescent dye 1 in channel C3 is not expected to be as severe. Similarly if another fluorescent dye with an emission maximum at a longer wavelength than that of fluorescent dye 2 and the limits of channel C2 were moved to match then the overlap from fluorescent dye 1 in C2 would be decreased. In the case where the fluorescent dyes are excited by different lasers with widely different excitation wavelengths, for example a red dye excited at 633 nm and emitting at 672 nm and a blue dye excited at 400 nm and emitting at 465 nm, the spectral overlap for these dyes will be non-existent.

The effect of spectral overlap on the measured data is shown in Figure 2. This plot shows the data measured from 10,000 commercially available microspheres (Luminex 64). These microspheres have been dyed with two fluorescent dyes in 64 different combinations of concentration. As can be seen from the plot, microspheres with low concentrations of both dyes result in almost circular clusters, this is a result of the variations in fluorescence intensities, which are greater than any overlap effect that is present. In contrast, microspheres with larger concentrations of fluorescent dye 1 are more elliptical and the axes of the ellipse are tilted with respect to the axes of the graph.

Spectral overlap is a well known problem in flow cytometry and other disciplines which use spectroscopic methods. In flow cytometry, "electronic compensation" can be used to reduce the effect of spectral overlap. However, electronic compensation (the subtracting out of cross talk components) is difficult to specify as the number of sensory channels and cross-talk interactions increases.

An outline of the process of histogram analysis is shown in Figure 3. Briefly, the process is started by collecting the fluorescence measurements of for the encoding channels of a number of empty microspheres 1. The number of microspheres measured will be determined by the complexity of the microsphere set and the sample size required to correctly estimate the clustering parameters for each cluster present. In the experimental work done to date, 10,000 microspheres were measured to classify 64 clusters in accordance with the Luminex protocol. Generally the number of microspheres required for data collection will be at least about 100 times the number of addresses.

The data is stored in an electronic format 2. If the data analysis computer is different from the data collection computer the information is transferred 3 to where it will be read by the analysis software. A histogram is built for the C1 measurements 4. In an alternative embodiment of the invention, the histogram analysis is independently performed for the fluorescent dyes for each laser, in which case each channel C1 will refer to the first encoding channel for a laser. This can be done because of the expected absence of spectral overlap between lasers. The results from these analyses are then combined to produce the full classification.

The limits of the peaks in each histogram, herein referred to as "bin limits" are found using a manual 5a, or a semi-automated 5b method. These methods are further described below. The bins defined in step 5a or 5b are used to segment the microspheres 6. Each segment is referred to as a bin. For each bin built in step 6, a histogram is built using the measurements of the next channel. The steps of 5, 6, and 7 are repeated for each channel, until all encoding channels have been analyzed. For channel C1 there will be a number of bins, for channel C2 there will be a set of bins for each C1 bin, these "bins of bins" can also be referred to as clusters.

Two methods for determining the bin limits may be used, 5a and 5b. In one, the user of the analysis (clustering) software sets a number of parameters, which are then used to identify the start and end of a bin. The alternative method is more automated (see "Ad Oculos Image Processing" by H. Bässmann and Ph. W. Besslich (Chapter 5, pages 46-51)).

As an example of histogram analysis, the histogram for the C1 data shown in Figure 2 is shown in Figure 4. A histogram is a standard method of displaying the data for one channel

in flow cytometry. Briefly the histogram is a plot of the count of microspheres that have a particular value of C1.

The first analysis method requires the user to specify a threshold value, a minimum "run length" and a minimum bin size. These terms are explained below, as required. The analysis  
5 iterates from the left of the histogram to the right. Starting at the left of the histogram, a flag is set to indicate that it is not inside a peak. Each microsphere count is compared with the threshold. The threshold is the minimum microsphere count required for the start or end of bin. For the data shown in Figure 4, the threshold was set to 0. If the count is above threshold, the current value of C1 is recorded as the start of a bin, and the flag indicating that it is inside a peak  
10 is appropriately set. While inside the peak the next count is checked successively. When a microsphere count below the threshold is found; a check is made that a number of the following C1 values are all below threshold. The number of C1 values required to remain below threshold is called the minimum "run length".

If the current C1 value meets the threshold and minimum "run length" requirements, the  
15 C1 value at the start of the bin is compared to the proposed end of bin value. If the difference is larger than the minimum bin size, then the current value of C1 is recorded as the end of bin. The flag indicating being within a peak is reset appropriately. Using the start and end bin values, a new bin entity is added to the end of a list of bins found. The binning steps are repeated, checking the microsphere count for each possible value of C1. When complete, a list of bins will  
20 have been built. This completes the definition of the bins (corresponding to step 5A in the flow chart). The semi-automated method will also result in a list of bin with a start of bin and end of bin limit (this corresponds to step 5B in the flow chart).

Step 6 in the flow chart is the assignment of microspheres to bins (for C1) or clusters for C2 and higher. The following explains the assignment of microspheres for the first iteration. The  
25 C1 value of each microsphere is compared with the start and end bin values of the bins in the list, starting with the first bin in the list:

(a) If the value of C1 for the microsphere is greater than the start of bin value and less than the end of bin value, then the microsphere is added to the list of microspheres in the bin and assignment of the next microsphere is started.

30 (b) If the test (a) fails, the test is repeated for the next bin in the list.

(c) Continue until the microsphere has been assigned to a bin or until no bins remain.

It is possible that some microspheres will not to be included in a bin, or a cluster when repeating for C2 and higher. These missed microspheres will be classified when a refinement of the classification is made, as described later.

35 When assigning microspheres to bins for C2 or higher there is a modification to the method outlined above. The modified step is performed before step (a) above. For each bin of

the previous channel, for example for each C1 bin when assigning to C2 bins, assign microspheres using bins derived from the C2 histogram for this C1 bin.

From the flow chart at step 7 a histogram is built for each bin (or cluster) made during step 6. Then at step 5A or 5B the bin limits are found for each of the peaks in those histograms.

- 5 The modified step described above indicates that only the microspheres used to build a histogram in step 7 are assigned to bins derived from that histogram at the next iteration of step 6.

To illustrate this point. One of the histograms of C2 for one C1 bin is shown in Figure 5. The C1 bin is from the right hand side of Figure 2. As can be seen, the peaks are not as well  
10 separated as those shown in Figure 4. When this data was analyzed using the manual method many combinations of threshold, minimum "run length" and minimum bin size were attempted before the correct number of clusters were obtained. In contrast when the method used for step 5B was used one set of parameters were used to determine the bins for both C1 and C2.

The computation time for this method will increase with increasing complexity of  
15 microsphere sets. The number of histograms that need to be analyzed will increase as  $O(\log N)$ . The time required to construct each histogram will increase as  $O(N)$ . Therefore the overall expected increase in computational time will be  $O(N \log N)$ . An advantage of this method over other clustering methods is that it results in a classification that is closer to the ideal, and therefore requires fewer iterations by k means, resulting in a better overall performance.

20 Once the microspheres are classified by assignment to a cluster, the clusters are characterized by calculating the cluster centers and the spread of each cluster. The cluster centers are taken to be the average channel values of the microspheres in the cluster, and the spread of a cluster is measured using the standard deviation channel values for the microspheres in the cluster.

25 Due to the spectral overlap of lower channels into higher channels, the clusters tend to be tilted. Each cluster can be analyzed using principle components analysis (PCA) to determine the orientation of the cluster. The principle components can be used for automatic compensation.

30 As mentioned earlier, not all microspheres will be assigned to a cluster. This can happen if the microsphere has a measurement that is between two bins for one of the channels. The cluster assignment can be refined using a standard clustering technique known as k means.

To perform k means the position of the centers of the clusters must be determined. This is done by calculating the average across the microspheres in a cluster for each channel. The process is outlined below:

- 35 A. For each cluster found.  
B. For each channel (C1, C2, ...).  
C. Calculate the average value of the channel for the microspheres in the cluster.

D. Repeat for all channels.

E. Repeat for all clusters.

Once the cluster centers have been found, the initial assignments are cleared and all microspheres are reassigned. The reassignment is performed by comparing the microsphere values for the channels used to define the clustering with the centers of all clusters. The microsphere is assigned to the cluster whose center is closest to the microsphere. The process for initializing k means is as follows:

A. Clear microsphere assignments.

B. For each microsphere in the data set:

10 Calculate the Euclidean distance between the microsphere and the cluster center. The Euclidean distance is a well-known mathematical quantity, which is defined as  $d = \sqrt{\sum (b_i - cc_i)^2}$ . Where d is the distance between a microsphere and a cluster center,  $b_i$  is the microsphere value for channel i and  $cc_i$  is the average value for channel i calculated for the cluster previously.

15

This initialization process ensures that every microsphere in the data set is assigned to a cluster. The cluster centers can be recalculated using the new assignments. The refinement process can be continued using the k means iterative process. This process as implemented in the present methods is as follows:

20 1. For each cluster identify the  $5^N - 1$  nearest clusters using the Euclidean distance between cluster centers, where N = number of channels used to define clusters, or the total number of fluorescent dyes used for classification.

2. For each cluster compare each microsphere in the cluster with the  $5^N - 1$  nearest neighboring clusters, by calculating the Euclidean distance between the microsphere and the centers of neighboring clusters. If a microsphere is closer to the center of a neighboring cluster than the center of the currently assigned cluster, then reassign the microsphere to the neighboring cluster.

3. Recalculate the cluster centers.

4. Count the number of reassignments made in step 3. If count is greater than zero, repeat from step 1.

30 The k means method is a standard method of cluster analysis and is known in the art. For example, see Massart *et al.* "Chemometrics: a textbook" pp 379 – 380. In the subject methods, however, the algorithm is optimized by comparing each microsphere with only the nearest neighboring clusters. This optimization reduces the computational time for complex microsphere sets, e.g. those comprising more than  $10^4$  addresses.

The effect of spectral overlap is to tilt clusters with higher concentrations of fluorescent dyes. In an optimal sorting strategy, this tilt must be accounted for. First the orientation of a cluster with respect to the axes is determined. This is achieved using a method known as Principle Components Analysis (PCA), which is a well-known technique (see, for example, 5 Massart *et al.*, *supra*. pp 403-407).

The Principle Components (PCs) of each cluster are calculated by first calculating the covariance matrix of the cluster using standard techniques. The covariance matrix is then diagonalized using "Jacobi Transformations" as described in "Numerical Recipes in Pascal" by William H. Press *et al.* The diagonalization of the matrix results in a set of eigenvalues, which 10 describe the significance of a PC and a set of eigenvectors that describe the orientation of the cluster. The eigenvalues and eigenvectors are sorted in descending order by eigenvalue.

At this point one has a complete characterization of all the clusters. The cluster centers can be used to determine the cluster position, and the PCs to determine how the clusters are oriented. The spread of the points around the centers may be calculated using either the 15 standard deviations in the encoding channel values or by transforming the microsphere positions into PC coordinates and calculating the standard deviations along the PCs.

This can be achieved using the following method:

For each cluster: convert each microsphere position from channel coordinates to PC coordinates by performing vector multiplication of the channel coordinates with each of the 20 eigenvectors in turn. The first eigenvector yielding the first PC coordinate and second eigenvector the second PC and so on. Calculate the standard deviation for each PC across all microspheres in the cluster. Repeat for each cluster in turn.

This characterization can be used to generate a number of geometrical shapes that mark the boundary demarcating the inclusion in, and exclusion from, a cluster. These include but are 25 not limited to, hyperellipsoids, hyperspheres, hypercubes, hypercylinders and cigar shapes.

When analyzing larger microsphere sets with more classification channels, automatic compensation may be performed prior to histogram analysis. The PCA may be applied to the complete data set, rather than on individual clusters. Histogram analysis may be performed on the data after a transform has been applied to remove the effect of spectral overlap.

30 Scaling up the complexity of the microspheres sets will lead to a number of problems. First will be the need to characterize more clusters. The optimization of the k means as described above will lead to better performance on high complexity sets, as compared to normal k means, because each microsphere is compared to only the closest neighboring clusters rather than to all clusters. As the complexity of the microsphere set increases this will result in more 35 dyes being used, and therefore more histograms to be analyzed. For example if a  $10^6$  microsphere set is prepared using 6 fluorescent dyes each with 10 levels, then  $1 + 10 + 100 +$

1,000 + 10,000 + 100,000 =  $1.11111 \times 10^5$  histograms will have to be analyzed using the method described above.

Due to the limitations of fluorescent dyes, with increasing complexity more lasers may be required to excite the dyes. As mentioned above, there will be no spectral overlap between fluorescent dyes excited by different lasers. If there is no spectral overlap between two neighboring classification channels, then they can be analyzed independently and the results combined later.

For example, a microsphere set of complexity  $10^4$  addresses is prepared, using four fluorescent dyes, excited by two lasers, using 10 levels for each dye. Fluorescent dyes 1 and 2 are excited by laser 1 and are measured in channels C1 and C2, while fluorescent dyes 3 and 4 are excited by laser 2 and are measured in channels C3 and C4. Furthermore, there is spectral overlap between channels C1 and C2; C3 and C4 but not between C2 and C3. Since there is no spectral overlap between C2 and C3 these can be analyzed independently. However, channel C2 must be analyzed after C1 and C4 after C3.

Analysis of this microsphere set is performed using two independent histogram analyses. The first analysis will analyze channels C1 and C2 using the methods shown in the flow chart, which results in a set of 100 clusters. Each of these clusters will include microspheres with all possible encodings of fluorescent dyes 3 and 4, but since there is no spectral overlap from these fluorescent dyes the encoding will have no effect on the cluster characteristics. The second analysis analyzes channels C3 and C4, using the method described in the flow chart on Figure 3. With C3 replacing C1 in steps 4 through to 7, this will result in another set of 100 clusters. Each of these clusters will contain microspheres with every possible encoding of fluorescent dyes 1 and 2. Microspheres can be assigned to one of the  $10^4$  possible clusters by forming the cross product of the two sets of clusters. This is achieved by segmenting every one of the 100 C1/C2 clusters using the C3/C4 cluster membership.

One of the goals of the classification methods is to be able to assign a unique identifier to each cluster. The identity of the cluster thus formed can be constructed as a number AB where A is a number between 0 and 99, and B is a number between 0 and 99. Each cluster will have a unique number between 00 to 9999. The value of A is determined by membership in a C1/C2 cluster (where each C1/C2 cluster can be numbered from 0 to 99; A = the number of the C1/C2 cluster that a microsphere has been assigned to). Similarly the value of B is determined by membership in a C3/C4 cluster (where each C3/C4 cluster is assigned a number from 0 to 99; B = the C3/C4 cluster that it has been assigned to). Therefore, each microsphere is assigned to one of the  $10^4$  possible clusters based on its membership in one C1/C2 cluster and one C3/C4 cluster.

The subject methods are less computationally demanding, since only  $1+10+1+10=22$  histograms were analyzed instead of  $1+10+100+1000=1111$  histograms which would have to be

analyzed if the channels were analyzed in the order C1/C2/C3/C4 to produce the same clusters. After the initial identification of the clusters and assignment of microspheres, k means can be used to refine the clusters as before.

- The association of the clusters to their sorting parameters is stored electronically.
- 5 Knowledge of the cluster centers, spread and orientation are used to determine sorting parameters that are sent to the flow sorter. These parameters determine the destination of each microsphere as it is sorted.

#### SORTING ASSIGNMENTS

- 10 Once a microsphere set has been classified, it can be used in synthesis reactions. A computer file assigning a unique oligonucleotide sequence to each encoded microsphere is created to control the flow sorting. The starting mixture of fluorescently encoded microspheres is then sorted to the outputs of the flow cytometer, depending upon which monomer will be attached to the nascent chain on the microsphere. The microspheres are then transferred to a
- 15 synthesis instrument and the appropriate coupling reaction is performed. These microspheres are then pooled and returned to the flow sorter and the sorting and synthesis process is repeated until synthesis is complete. Libraries of  $4^N$  components are prepared using only N sorting steps and 4N coupling reactions

- A goal of the subject methods is to assign sequences to microsphere clusters and enable
- 20 sorting parameters to be generated for transfer to the sorter, and to minimize the number of microspheres that have an incorrect sequence synthesized on them.

- A microsphere may have an incorrect sequence synthesized if during one or more sorting cycles it was incorrectly sorted to the wrong output. There are two possible causes for the error. The first is that the microsphere was incorrectly classified due to random variations in dye
- 25 concentrations that occur during the encoding process; random fluctuations in the measured fluorescence and errors in the classification software. The second cause for incorrect sorting is random fluctuations in the sorting mechanism that misdirect a microsphere.

- When an error of the first type occurs, a microsphere is most likely to be misclassified as belonging to one of its neighboring clusters. When an error of the second type occurs, either the
- 30 microsphere is lost completely or it is sorted into a neighboring output container. An optimal assignment of sequence to microsphere clusters accounts for both types of errors. Each source of error can be characterized by measuring the probability that a microsphere that has been assigned to one output will be sorted into an output, either to the intended output or a different output. These probabilities can be tabulated as the following example shows for a four-way
- 35 sorter.



	Output 1	Output 2	Output 3	Output 4
Output 1	0.98	0.01	0.005	0.0005
Output 2	0.0075	0.98	0.0075	0.003
Output 3	0.001	0.01	0.97	0.01
Output 4	0.0005	0.005	0.01	0.90

The probability of a microsphere that is intended for a particular output being sorted to an actual output is tabulated in each column. The column headers are the intended outputs. The most probable actual output is the intended output, but some microspheres will be sorted into the neighboring outputs. Not all microspheres are collected, therefore the sum of these probabilities is less than one. This is an example of the second type of error. The probabilities in the table are determined empirically using optimized flow setting, e.g. flow rate, microsphere concentration, plate voltages, etc., for the sorting errors; and for each microsphere set for the classification errors.

These tables are used to optimize the assignment of sequences, using optimization methods known in the art. The optimization will build a mathematical model of the sorting process, using the tabulated probabilities to predict the fraction of microspheres incorrectly sorted for a particular assignment of sequence to microsphere cluster and assignment of nucleotide to sorter output. Monte Carlo methods may be used in the optimization. An optimization method is then used to produce an assignment of sequence to microspheres and nucleotide to sorter output. Suitable methods include, but are not limited to, genetic algorithms, simplex optimization, and the like, as known in the art. For example, see Heitkoetter and Beasley, eds. (1999) "The HitchHiker's Guide to Evolutionary Computation: A list of Frequently Asked Questions (FAQ)", USENET: comp.ai.genetic. Available via anonymous FTP from rtfm.mit.edu/pub/usenet/news.answers/ai-faq/genetic/. A description of Simplex Optimization can be found in many texts including Massart *et al*, *supra*.

Having characterized the clusters present in the data and assigned sequences to those clusters, this information must be encoded into a format suitable for use with the sorting instrument and transferred to a controller for the sorter. The format of the encoding will depend on how the sorting decision making is performed.

There are several methods by which the decision on where a microsphere is to be sorted can be made. In one method the similarity of a microsphere to be sorted is computed for each of the clusters. The microsphere is then sorted according to which cluster it is most similar to. The steps to make this decision are: to calculate distance between measured values for the microsphere and each cluster center. Assign the microsphere to closest cluster, using the same process as was used to assign microspheres to clusters for k means. Calculate the PC

coordinate for the microsphere. Compare the microsphere PC coordinates with the cluster spread along PCs.

This method assigns the microsphere to the closest cluster, then uses information about the Principle Component and the spread of the cluster along the PC axes to determine if the microsphere is close enough to the cluster center to be considered a part of the cluster. If the microsphere is accepted as part of the cluster it is sorted to the output assigned to the cluster, otherwise it is sorted to waste. The boundary between accepting a microsphere in a cluster and rejecting it can take a number of different shapes depending on how the comparison between microsphere coordinates and the spread is made. These shapes will include hyperspheres, hyperellipsoids and others.

This method of sorting designation is computationally intensive. As the complexity of the microsphere set increases, the microsphere value are compared to more clusters, which slows the decision making process. In a preferred method, high-speed parallel digital signal processors (DSP) are used to compare the microsphere with a subset of the clusters. If the microsphere can be assigned to a cluster a signal will be sent to sort the microsphere, otherwise a signal to send the microsphere to waste is generated. As the complexity of the microsphere sets increases, the number of processors in the system will also increase.

One design for parallel DSP sorting designation is shown in Figure 6. In Figure 6a is shown how a 64 microsphere set is divided into 4 subsets of 16 clusters each. In Figure 6b, a host computer sends the cluster parameters to the four processors using high-speed serial communications channels. Each DSP processor stores these parameters in local memory (M2) for later comparison with the microsphere data. The host computer receives the microsphere data on an instrument interface. This data is distributed to all processors simultaneously by writing a copy into an area of shared memory (M1) present in each processor module. Once this is complete, a signal is sent to all processors indicating new data is present. Each processor then independently compares the data with the cluster centers stored in local memory. If the microsphere is within a cluster only one processor will obtain a match, this processor signals the sort destination using a sort signal bus. If no match is found, then no sort signal is generated and the microsphere goes to waste. Since there is no inter-processor communication during the decision making process, this design scales linearly with the number of processors. That is, if one processor takes  $t$  microseconds ( $\mu s$ ) to make a decision for  $k$  clusters, then  $N$  processors each with a subset of  $k/N$  clusters will take  $t/N$   $\mu s$  to make the same decision.

A number of methods have been described in the art for multiple parameter sorting, increasing processing speeds and accuracy of flow cytometer cell sorters. For example, systems have been proposed for electronics modularization to process as many as eight input parameters for sorting cells (Hiebert *et al.* (1980) *Cytometry* 1:337, 1980). A system for correlating multiparameter data for each cell is described in Parson *et al.* (1985) *Cytometry* 64:388; and a

system for parallel processing a signal from a large number of detectors by van den Engh (1989) Cytometry 10:282.

#### LOOK UP TABLES

5        Flow sorters often use look up tables (LUTs) to determine where a particle is sorted. The LUT is addressed using the measured fluorescence and the output as a signal that selects the destination of the particle. In currently used sorters the fluorescence is usually digitized as a 10-bit number (in the range from 0 to 1023). Therefore a LUT for one channel requires only 1024 entries. LUTs have the advantage of high speed but are limited to setting ranges between which  
10        particles are selected.

      A lookup table can be constructed from an electronic memory device that is programmed with the desired results. An electronic memory device has a set of inputs known as addressing lines and a set of outputs known as data lines. In the application of LUTs to flow cytometry, the memory device is programmed with the sorter outputs for the microspheres. The digitized  
15        measurements from the fluorescence channels are placed on the address lines, this address is decoded by the internal circuitry of the memory device and destination of the microspheres is output on the data lines. In Figure 7 a simple lookup table is shown, with only 11 addresses, two sorting destinations and a waste output. The eleven addresses can be represented using a 4 bit binary number, so only 4 address lines are required. The three possible sorting options can  
20        be represented using a 2 bit number so only two data lines are shown. In general the microsphere data is routed from the instrument interface, perhaps by means of a host computer, to the address lines of the LUT and the data lines are used to generate an appropriate sorting signal.

      A variation on the lookup table is a bitmap. The bitmap is a lookup table using two  
25        fluorescent channels to address it. The address lines are grouped into two logical groups and the two digitized measurements access the LUT using one group of address lines each. A diagram of how a bitmap is implemented using a memory device is shown in Figure 8. This figure shows a simple bitmap for two encoding channels (C1 and C2) which can have values from 0 to 3 (for a typical flow cytometry instrument the range is at least 0 to 1023). This results  
30        in 16 combinations of C1 and C2. The bitmap shows for every combination of C1 and C2, where the microsphere or particle is to be sorted (Figure 8A). Figure 8B shows how the two dimensional bitmap can be converted to a lookup table; while Figure 8C shows how a memory device programmed with the LUT from Figure 8B is addressed to implement the bitmap of Figure 8A. The strategy of a bitmap can be generalized to any number of fluorescent channels, limited  
35        only by the size of memory devices.

      For highly complex arrays, the look up table is modified to meet the memory requirements. Methods to reduce the amount of memory space required to represent the data

include, without limitation, hierarchical LUT, using memory management units, sparse arrays and hash tables or combinations of these.

Hierarchical look up tables are a method of reducing the amount of memory required by exploiting the properties of spectral overlap, where there is a lack of spectral overlap for  
5 fluorescent dyes excited with different lasers. Fluorescent dyes excited with different lasers can also be sorted by combining the output of two LUTs (or bitmaps). The first LUT will use channels C1 and C2 to address a memory device, the contents of the memory having been programmed with the two digit cluster identifiers (00 to 99). The second LUT will use channels C3 and C4 to address a second memory device, which has been programmed with two digit cluster identifiers  
10 (00 to 99). The combination of these two outputs will produce a code from 00 to 9999. The outputs from these two LUTs are used to address a third memory device (the second layer of a hierarchical LUT) which has been programmed the sorting destinations. The memory requirements for this system is substantially less than that required for a four dimensional bitmap.

The construction of a hierarchical LUT follows the histogram analysis strategy that was  
15 used in the initial classification. If the histogram analysis is divided into a number of independent analyses, then the first level of the hierarchical LUT will consist of that same number of LUTs. The channels that are analyzed together, are combined to address a LUT.

An example is shown in Figure 9. In this example a microsphere set is used with 16 clusters, using 4 fluorescent dyes at two concentrations and using two lasers. The bitmaps are  
20 shown in Figure 9A. The assignment of 16 sequences to the microspheres is shown in the table. The overall cluster ID is formed by setting the first digit to the C3/C4 cluster ID and the second digit to the C1/C2 cluster ID. Figure 9B shows the bitmap required for synthesizing the first base. The hardware implementation of this system is shown in Figure 9C. To perform the sorting to synthesize the second and subsequent bases only the second bitmap (shown in Figure 9B)  
25 needs to be reprogrammed.

Usually the number of first level LUTs will equal the number of lasers used to excite the fluorescent dyes. The number of channels used to address a first level LUT will equal the number of fluorescent dyes excited by the laser associated with the LUT. One implementation of a  $10^6$  microsphere set will use two lasers, e.g. red and blue, with two fluorescent dyes excited  
30 by the red laser and four fluorescent dyes excited by the blue laser. This implementation may be combined with another method of reducing memory space.

A Memory management Unit (MMU) is an electronic device present in almost all computers, which allows the computer to access more memory than is installed in the computer. Its operation is well known in the art, and has been described in a number of texts. The basic  
35 operation of the device is to translate a "virtual" address into a physical address.

In the present methods, the virtual address is formed using the digitized measurements from the fluorescent channels. The physical address is the address of the memory device used

to store the LUT. The methods take advantage of the fact that not every possible virtual address is required to store information on the clusters, only the addresses within clusters need to be used. An address is considered to be "within" a cluster, if the channel data values used to construct the address would be classified as being inside the cluster boundary. Otherwise the  
5 address is regarded as being between clusters. Information is not required about the space between clusters, since any microsphere which gives a measurement in this space will be sorted to waste.

The LUT will only contain information about the clusters. The MMU will compare the virtual address formed from the digitized values for a microsphere, and if the MMU can translate  
10 the virtual address to a physical address, this physical address will be used to lookup the cluster information. In the case of a first level hierarchical LUT this would be the cluster identifier, or part thereof. For a nonhierarchical LUT this would be sorting destination. If the MMU cannot translate the virtual address, a signal is generated that indicates the microsphere should be sorted to waste. Using an MMU a four channel bitmap is practical.

15 Sparse arrays and hash tables are standard software methods for storing large arrays of data when only a small fraction of the array is required. The representation of the clusters is an example of sparse data. The methods known in the art and described in texts. These methods store the data in a nonlinear array, and use a computational algorithm to translate the required address into a physical address before retrieving the data. Using parallel DSP systems, the  
20 computation is performed in the time required for efficient sorting.

#### Sorting Parameters

The software allows the user to specify a set of oligomer sequences to be synthesized. Each defined oligomeric sequence is associated with a distinct microsphere address. During  
25 the synthesis and sorting procedure, the clusters of microspheres are sorted into groups for reaction for monomers, then usually combined and resorted, such that in the end each cluster has a defined oligomeric sequence. The instructions for the set of sorting and synthesis reactions comprises a series of sorting parameter sets, where one set of sorting parameters is required per monomer. The set of sorting parameters are sent to the flow sorter during the  
30 synthesis process.

The methods of the present invention may used to create very large liquid arrays having microsphere sets with greater than  $10^4$  distinct addresses, and which may be greater than  $10^6$  different addresses. In such cases it may be necessary to use a microsphere set that is much larger than the number of different oligomer sequences. Where the library requires most or  
35 substantially all of the microsphere set, it is herein designated as a dense array. Applications that require only a few of the microspheres in the set are herein designated as a sparse array. The strategies for assigning oligomers to microspheres is different for these two cases.

The software generates an optimal assignment of oligomers to microsphere addresses, which will minimize the number of sequence errors due to microspheres sorted incorrectly. During the synthesis of a library, generally a plurality of microsphere clusters are sorted into separate groups for synthesis reactions. After coupling of the monomer, the group is then either  
5 split into subgroups during a second sorting process, or combined with the other groups and resorted. The sorting groups for a microsphere cluster may be designated by the round of synthesis (round 1 for first residue, round 2 for second residue, etc.) and by the monomer for coupling.

For example, where the oligomer is a polynucleotide, the first sorting group may be for  
10 coupling an adenosine, hence G<sub>1</sub>A; the second sorting group for thymidine coupling (G<sub>2</sub>T) and the third round for coupling a cytosine (G<sub>3</sub>C), which provides the sequence ATC. A different polynucleotide in the library, having the sequence ATG, would have the sorting group string G<sub>1</sub>A;G<sub>2</sub>T;G<sub>3</sub>G. The microsphere clusters for these two sequences would be sorted into one reaction group in the first two rounds, but would be split in the third sorting group.

15 For dense arrays, where the number of oligomers is equal to or nearly equal to the number of microspheres in the set, the goal is to assign oligomers with similar sequence to microspheres with similar encoding. This will mean that if a microsphere is incorrectly identified as belonging to a neighboring cluster it will still be sorted correctly if the required base is the same for both clusters. Oligomers with similar sequence may be assigned to clusters that are  
20 close together.

For sparse arrays, where the number of oligomers is very much less than the number of microspheres in the set) the goal is to assign oligomers with similar sequence to microspheres with less similar encodings and relax the sorting parameters. This arrangement will reduce the number of sorting errors. Alternatively, the effective size of the microsphere set is reduced by  
25 defining "superclusters". These are clusters of neighboring clusters, which result in more than one address being used per oligomer. The sorting parameters are a combination of the clusters that make the supercluster. The effective size of the microsphere set can also be reduced by reducing the dimensionality, that is the coding of one or more dyes is ignored. This effectively creates superclusters by collapsing one or more columns of clusters.

30 General purpose optimization methods, which include genetic algorithms and simplex optimization, can be applied to assignment of oligomer sequence to microspheres. Genetic algorithms start with a population of possible solutions and test each solution to determine if it is successful. The test uses a model that predicts the number of errors in sequence based on the microsphere assignment and a probability of incorrectly identifying a microsphere, e.g. using  
35 a Monte Carlo simulation. The solutions that result in the fewest errors will survive and be used to produce the next generation of solutions using software equivalent to point mutation and recombination.

The simplex optimization uses only a single possible solution, and attempts to improve successively that solution by making small changes in the microsphere assignments and testing if the solution has improved. If it does, then another larger change would be made. If it does not, then a small "backward step" is made. This process is iterated until the solution cannot be improved on.

Once an optimal assignment of oligomer sequence to microsphere has been made the sorting parameters for addition of each base are generated. The sorting parameters for each base are a combination of the sorting parameters of the microsphere clusters requiring addition of that base.

#### SYNTHETIC REACTIONS

The sorted microspheres of the invention are used to covalently attach a monomer that provides the starting point for the solid phase synthesis of a compound, usually an oligomeric compound. Conventional formation of an oligomer by stepwise addition of monomers to the microsphere may be performed. Alternatively, a monomer bound to microspheres may be further divided into groups and then chemically modified by introduction of substituents to form a series of analogs of the starting monomer. Oligomers of interest include oligopeptides, oligonucleotides, oligosaccharides, oligomers of peptide mimetics such as oligopeptoids, and the like. Conventional reagents and methods for making oligopeptides, oligopeptoids, oligonucleotides, and the like, can be used.

In combinatorial processes, the materials and techniques now used in combinatorial chemical techniques are known in the art and discussed, see for examples, Houghten (1985) *Proc. Natl. Acad. Sci. USA* 82:5131-5135; Geysen *et al.* (1984) *Proc. Natl. Acad. Sci. USA* 81:3998-4002; Pirrung *et al.* (1995), *J. Am. Chem. Soc.* 117:1240-1245; Smith *et al.* (1994) *BioMed Chem. Lett.* 4:2821-2824; Beaucage *et al.* (1981) *Tetrahedron Lett.* 22:1859-62, and Itakura *et al.* (1975) *J. Biol. Chem.* 250:4592 (1975).

After covalent attachment, the protected groups present on the nascent chain or substrate are deprotected using cleavage reagents appropriate to the selected protecting groups. The reaction is then initiated by adding a monomer, adding or deleting a substituent, *etc.* It is common in solid phase synthesis of oligomers, after the deprotection of the anchored molecule, to add reactive monomer to the reaction vessel. Such monomers usually comprise a protected moiety and a reactive moiety, *e.g.* an Fmoc protected amino acid. The reaction is allowed to proceed to completion, followed by washing steps, blocking steps, *etc.* as known in the art. As previously described, the synthesis will utilize a "split/mix" approach, wherein after every monomer addition, the contents of the reaction vessels are alternatively divided and mixed in a way that provides for a diverse set of ligands (see Pirrung *et al.*, *supra.*)

The distinct oligomers in the library provided may be screened for activity, e.g. by screening individual sublibraries containing mixtures of distinct oligomers, identifying active sublibraries, and then determining the oligomeric compounds of interest by generating different sublibraries and cross-correlating the results obtained.

5       References describing construction of small organic molecule libraries include: Thompson et al., *Chem. Rev.* 96:555-600 (1996); Gallop et al., *J. Med. Chem.* 37:1233-1251 (1994); and Gordon et al., *J. Med. Chem.* 37:1385-1401 (1994). A reference related to mimotopes and describing the construction of peptides on solid supports is U.S. Patent No. 4,708,871 to Geysen et al., while other references generally describing construction of peptoid  
10   libraries include Bartlett et al., PCT Publication No. WO91/19735, and Zuckermann et al., PCT Publication No. WO94/06451. References describing screening of compounds and determination of sequences include U.S. Patent Nos. 4,833,092 to Geysen et al., 5,194,392 to Geysen et al., 5,573,905 to Lerner et al., and 5,585,277 to Bowie et al.

Use of a cleavable linker in this system allows the synthesis of large numbers of different  
15   oligonucleotides for use as solution phase primers or probes. Base deprotection and cleavage can be carried out on the entire set in cases where multiplexed pools of probes or primers are required, e.g. for use in multiplex PCR or in specific priming during reverse transcription of mRNA. Alternatively, the system provides a means to synthesize small amounts of each individual oligonucleotide, if each microsphere address is sorted into a separate tube prior to  
20   deprotection and cleavage.

#### USE OF LIQUID ARRAYS

The methods of the present invention are used for the creation of libraries of oligomers coupled to addressed microspheres. The oligomers may be cleaved from the microspheres and  
25   used in a conventional method; or may be retained on the microspheres and utilized in assays that exploit the address features for analysis. Methods of screening for small molecules may be performed, as is known in the art. Peptide libraries find use in binding studies, as epitopes for immunological studies, for studies of biological activity *in vivo* or *in vitro*, and the like.

Where the oligomers are oligonucleotides, the arrays find use in the areas of gene re-  
30   sequencing, polymorphism typing, and gene expression quantification. In a typical assay, a sample comprising a potential binding partner for one or more of the oligomers in the arrays is labeled with a fluorescent detectable label. The labeled sample is then combined with the array of microsphere conjugated oligomers. After binding is complete, the unbound sample may be washed away or otherwise removed.

35       Scoring (genotyping) a known single nucleotide polymorphism (SNP) in genomic DNA involves extraction of genomic DNA from a suitable source, e.g. buccal swab, whole blood, tumor



biopsy) and scoring the DNA for presence/absence of each allele to ascertain the genotype. In human genomic DNA, each allele is present as 0, 1, or 2 copies per genome. If the polymorphism being typed is subclonal (i.e. in the cases where tumors are being analyzed, genetic instability may alter the number of copies in some cells) then the allelic ratio can vary continuously between 0% and 100%.

Measurement of gene expression levels by analyzing mRNA levels provides another potentially important diagnostic method. Conventional methods for gene expression quantitation include Northern blots, RNase protection assay and quantitative RT-PCR. The assay is performed as described above, but quantitation of labeled probe is determined. Such measurements are usually normalized to a control sequence, e.g. housekeeping genes such as actin, tubulins, etc.

For gene re-sequencing, DNA is extracted and purified using an appropriate method. The DNA sequence of interest is amplified by PCR, with incorporation of a fluorescent label if required. Oligonucleotide probes synthesized on the encoded microspheres may be used to test the PCR amplicon at each base for the presence/absence of the wild-type base or a polymorphic base. Both strands of each PCR amplicon will be tested to improve data quality. The method for testing can be hybridization, or hybridization with enzymatic modification. Where there are known polymorphisms in the gene sequence, alternate panels will be developed in order to accurately scan the bases close to the known polymorphism.

A similar approach for typing of known mutations will be used, except the initial PCR may be a multiplex PCR, and not all bases within each PCR amplicon are tested. For gene expression quantification, mRNA will be extracted from the patient sample. Oligonucleotide probes will be synthesized on the microspheres to assay for the amount of each mRNA species present in the sample - these probes will be designed to be approximately 20-mers, and many probes per mRNA will be used to improve data quality and to assay for alternate splicing. Measurement of the expression levels of key genes is expected to provide high quality prognostic and best course of treatment information in cancer treatment.

#### EXPERIMENTAL

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the subject invention, and are not intended to limit the scope of what is regarded as the invention. Efforts have been made to ensure accuracy with respect to the numbers used (e.g. amounts, temperature, concentrations, etc.) but some experimental errors and deviations should be allowed for. Unless otherwise indicated, parts are parts by weight, molecular weight is average molecular weight, temperature is in degrees centigrade; and pressure is at or near atmospheric.

### Example 1

#### Synthesis by Sorting

Synthesis after one round of sorting was performed on a four microsphere (55% DVB polystyrene) set, demonstrating the *in situ* synthesis of oligonucleotides on microspheres labeled with fluorescent dyes, and performing synthesis after flow sorting based on differences in fluorescence emission intensities of each dye encoded microsphere. The experiment involves synthesis of a common 8-mer on all 4 microsphere sets, followed by flow sorting and continuation of synthesis of unique sequences (4-mer) on each of 4 microsphere set followed by another common 8-mer to complete the 20-mer sequence. The sequence on each microsphere set is then detected by hybridization with a fluorescently labeled oligonucleotide sequence (a probe which is complementary to the sequence on the microsphere) followed by flow cytometric detection of enhanced emission intensity.

a. *Derivatization of microspheres.* Samples of polystyrene (55% DVB/20%GMA, 8.8  $\mu$ ) microspheres with -NH<sub>2</sub> functional groups were a gift from Dyno Particles (Norway) or purchased from Bangs Laboratories. Dry microspheres were recovered from the suspension by filtering through a Whatman type 42 filter paper followed by washes with water, methanol acetonitrile and dichloromethane. The surface loading of amino functional groups was estimated to be 4.41  $\mu$ mol/g, determined by the method described in Reddy and Voelter (1988) J. Peptide Protein Research 31:345-348. The linker was attached to the surface amine sites of microspheres by direct coupling of carboxyl group on linker with aliphatic amine sites on particles as shown in figure 12.

DMT loaded and carboxyl derivatized C-12 linker [DMTO-(CH<sub>2</sub>)<sub>11</sub>COOH] was dissolved in dry acetonitrile and combined with 0.5 g of microspheres and diisopropylethylamine. HBTU and dimethylaminopyridine were dissolved separately in acetonitrile. This solution was injected (Hamilton gas tight syringe) into the microsphere suspension, vortexed and allowed to react for 30 minutes.

The microspheres were filtered, washed and allowed to air dry. The DMT loading of the microspheres was estimated to be 1.80  $\mu$ mol/g by measuring the absorbance of the released trityl ions at 503 nm (the amount of linker on microspheres can be controlled by varying the time of coupling reaction).

#### *b. Synthesis of Oligonucleotides on C-12 linked Supports*

*Instrument:* PE Biosystems Model 394 (4 column, 8 base) Auto Synthesizer.

*Column preparation:* 10-15 mg of C-12 linked microsphere support was weighed directly in to an empty synthesis column (1  $\mu$ mol columns from PE Biosystems) sealed on one end with 5-10 micron pore sized Zitex filter from Norton Plastics. The other end was then sealed and capped (Aluminum cap) both ends using a crimper tool.

*Capping of underivatized surface sites on dye loaded or C-12 linked microspheres:* The capping was done manually on the synthesizer by exposing microspheres to capping solution.

The microspheres were then washed by acetonitrile flow through the column. As an alternate a CAP BEGIN program was inserted which allows multiple capping/washing steps prior to synthesis.

*Synthesis cycle:* The 0.2  $\mu$ mol oligonucleotide synthesis cycle was modified by increasing washing and reagent addition times to compensate for lower flow rates through column. Flow rates were measured as gms/30 sec of reagent flow through column 1.

*Reagents:* Standard PE Biosystems reagents. Protected phosphoramidites used: A<sup>bz</sup>, G<sup>DMF</sup>, C<sup>bz</sup> (PE Biosystems).

#### *Oligonucleotide Synthesis Product Analysis*

*Trityl analysis:* The coupling efficiency was measured by monitoring the absorbance (503 nm) of released trityl ions following detritylation step. The trityl output was collected by fraction collector. The solution was allowed to evaporate to dryness in a fume hood. TCA/DCM solution was added to each tube and the weight of solution was measured. The absorbance was measured using a spectrophotometer. The absorbance readings were corrected for discrepancy in volumes due to solvent evaporation. Trityl data for 20-mer synthesis is shown in Tables 1 and 2.

Table 1 contains trityl absorbance data for a 20-mer synthesized on PE Biosystems 40 nmol polystyrene column using 40 nmol CE synthesis cycle. Absorbance values were recorded at 503 nm after dissolving evaporated residue of each fraction collected in detritylating reagent to a volume of 2 ml. The absorbance is corrected for discrepancies in volume due to evaporation of solvent.

Table 1

FRACTION	*A:nm
1	1.76
2	
3	2.03
4	2.04
5	1.98
6	2
7	1.96
8	1.91
9	1.95
10	1.89
11	1.88
12	1.84
13	1.82
14	1.84
15	1.83
16	1.78
17	1.77
18	1.86
19	1.79
20	1.76
21	1.75

Table 2 contains trityl absorbance data for 20-mer synthesized on 8.8 micron polystyrene microspheres (55 % DVB, with C-12-ODMT linker, 20 mg) modified by chemical phosphorylating phosphoramidite cleavable linker. Absorbance values were recorded at 503 nm after dissolving evaporated residue of each fraction collected, in detritylating reagent to a volume of 1.5 ml. The absorbances are corrected for discrepancies in volume due to evaporation of solvent. Fraction 1 corresponds to DMT off C-12 linker and fraction 2 corresponds to DMT off cleavable phosphorylating phosphoramidite.

Table 2

FRACTION	*A: nm
1	1.25
2	1.1
3	1.14
4	
5	1.07
6	1.03
7	0.82
8	0.79
9	0.84
10	0.88
11	0.73
12	0.73
13	0.73
14	0.69
15	0.66
16	0.61
17	0.66
18	0.64
19	0.57
20	0.54
21	0.61
22	0.6

*CE analysis:* A cleavable linker was used for cleavage and analysis by capillary electrophoresis. A 20-mer synthesized on 8.8 micron, 55% DVB polystyrene microsphere supports gave comparable results with same 20-mer sequence synthesized on standard 40 nmol (PE Biosystems) columns (Figure 13). Average stepwise yields of 99 % and 98.3 % were calculated for 20-mer synthesis on C-12+cleavable linker and 40 nmol PE Biosystems polystyrene columns respectively. No significant difference in product quality was observed for synthesis without C-12 linker attached.

Figure 13 shows capillary electrophorograms of a) 20-mer sequence synthesized on 8.8 micron polystyrene (55 % DVB crosslinked) microspheres; b) 20-mer sequence synthesized on PE Biosystems 40 nmol polystyrene support. Sequence synthesized; (SEQ ID NO:1) 5'>AGCT AGCT TTTT AGCT AGCT<3'. The products were simultaneously cleaved from support and deprotected in ammonium hydroxide (55° C, 16 h). A cleavable linker, [2[2-(4,4'-Dimethoxytrityloxy)ethylsulfonyl]ethyl-(2-cyanoethyl)-(N,N-diisopropyl)-phosphoramidite]] was linked to surface amino sites of polystyrene microspheres before synthesis of sequence.

*Mass spectroscopic analysis (MALDI):* 20-mer oligonucleotide synthesized on polystyrene (8.8 µ, 55 % DVB) was sent for MALDI analysis after cleavage and deprotection by ammonia (product was not purified). Expected mass = 6072; observed mass = 6073 (mass includes 3' phosphorylation from cleavable chemical phosphorylating reagent). As a control experiment the same sequence was synthesized on 40 nmol polystyrene support from PE Biosystems; expected

mass = 5990; observed mass = 5993. Mass spectroscopic analysis was performed by Mass Consortium in San Diego. Sequence synthesized: (SEQ ID NO:2) 5'> ATCCCCAACAGACCACTGCTC<3' DMT off.

5        *Fluorescent labeling by covalent attachment of Bodipy-TMR:* Fluorescently encoded microspheres tolerant to organic synthesis conditions were generated for protocol development purposes. The succinimidyl ester functional group of the dye (Bodipy TMR, Abs/Em = 542/ 574 nm) was coupled directly to the free surface aliphatic amine sites of C-12 (ODMT) linked microspheres in acetonitrile (~1.80 µmol/g linker sites) to form a carboxamide bond. The amount  
10 of dye attached to the surface of microspheres was controlled by varying the concentration of the dye in equilibration with the surface amine sites. The amount of C-12 linker (oligonucleotide sites) on the microspheres corresponds to ~ 40 % of total available sites, thus leaving about 60 % of sites available for dye binding.

15        Succinimidyl ester linked Bodipy TMR (molecular probes, D-6117, 1mg) was dissolved in dry acetonitrile. Three vials of C-12 linked microspheres was suspended in dry acetonitrile and reacted with 10, 40 and 100 µl of the dye solution and subjected to ultrasonication in a water bath for 60, 70 and 80 minutes respectively. The dye loaded microspheres were filtered and washed with acetonitrile. The background fluorescence emission from unlabeled C-12 linked  
20 microspheres was used as the fourth labeled microsphere set.

The following nomenclature is used to refer to above synthesized microspheres.

unlabelled microspheres	:	Microsphere 1
10 µl Bodipy stock /ml	:	Microsphere 2
25 40 µl Bodipy stock /ml	:	Microsphere 3
100 µl Bodipy stock/ml	:	Microsphere 4

To demonstrate the stability of covalently attached Bodipy dye to oligonucleotide synthesis conditions, histograms (FL2) were recorded before and after oligonucleotide  
30 synthesis/deprotection (Figure 14). Intensities for microspheres 2-4 remained unchanged. Increased intensity on microsphere 1 was due to added background emission from the synthetic process (reagents). Intensities were not changed significantly enough to hinder sorting experiment.

Figure 14 shows the effect of oligonucleotide synthesis conditions on a mixture of  
35 microsphere sets 1, 2, 3, 4; FL2 (orange) histograms of a mixture, before oligomer synthesis

treatment; b) after subjecting microspheres to 20-mer synthesis. c. Synthesis of 8-mer before sorting.

The four sets of fluorescently labeled microspheres were mixed together and inserted into a synthesis column. The sequence (SEQ ID NO:3) 5'> TCGA TCGA TTTT <3' was synthesized  
5 (DMT removed after last base addition) on microspheres 1,2,3 and 4.

*Sorting:* The microspheres were suspended and divided into two portions. Microspheres 1 and 2 were sorted from the first portion and microspheres 3 and 4 were sorted from the second portion. The sorted microspheres were collected in PBS buffer. The sorting was performed using  
10 a Becton Dickinson sorting flow cytometer. The histogram of mixed microspheres (before sorting) and histograms for each sort (sort 1, 2, 3, 4) are shown in Figure 15.

Figure 15: Sorting mixture of microspheres: FL2 (orange) histograms of; a) mixture of microspheres (1,2,3,4) before sorting; b) after sorting each component of mixture based on intensity of FL2 (orange) emission (sort 1, sort 2, sort 3, sort 4).

15

*Synthesis and deprotection after sorting:*

*Synthesis:* The sorted microsphere suspension was transferred into a syringe attached to the open end of a synthesis column. A Zitex (5-10 micron) filter was placed on the other end and sealed. The suspension was filtered by pushing plunger to apply mild pressure. The open  
20 end of the column was sealed (after inserting the filter) and inserted into oligonucleotide synthesizer.

Sorts 1,2,3 and 4 were transferred into columns 1,2,3,4 respectively and the following sequences were synthesized.

Microsphere 1: (SEQ ID NO:4) 5'> TCGA TCGA AAAA <3'  
25 Microsphere 2: (SEQ ID NO:5) 5'> TCGA TCGA GGGG <3'  
Microsphere 3: (SEQ ID NO:6) 5'> TCGA TCGA CCCC <3'  
Microsphere 4: (SEQ ID NO:7) 5'> TCGA TCGA TTTT <3'

*Deprotection:* The microspheres after oligonucleotide synthesis were dried, transferred  
30 into a microcentrifuge tube and treated with concentrated ammonia. The excess ammonia was decanted off and the microspheres washed. Deprotection for sorted microspheres were performed directly on synthesizer. The microspheres were recovered into synthesis columns followed by vortexing (to extract microspheres attached to filter and column).

35 *Detection of microsphere type (sequence) by hybridization:*

Microsphere 1: 5'> (SEQ ID NO:8) TCGA TCGA AAAA TCGA TCGA <3'

Microsphere 2: 5'> (SEQ ID NO:9) TCGA TCGA GGGG TCGA TCGA <3'

Microsphere 3: 5'> (SEQ ID NO:10) TCGA TCGA CCCC TCGA TCGA <3'

Microsphere 4: 5'> (SEQ ID NO:11) TCGA TCGA TTTT TCGA TCGA <3'

5 Probe 1: 5'> (SEQ ID NO:12) TCGA TCGA TTTT TCGA TCGA F<3'

Probe 2: 5'> (SEQ ID NO:13) TCGA TCGA CCCC TCGA TCGA F<3'

Probe 3: 5'> (SEQ ID NO:14) TCGA TCGA GGGG TCGA TCGA F <3'

Probe 4: 5'> (SEQ ID NO:15) TCGA TCGA AAAA TCGA TCGA F<3'

10 F= Fluorescein amidite

*Discrimination of complementary and non-complementary oligonucleotide hybridization to oligonucleotides synthesized on microspheres:* Microspheres 1,2,3 and 4 were mixed together and divided into 5 portions. Probes 1-4 (complementary sequences labeled with green fluorescein dye) were added to each tube (probe 1 to tube 1, probe 2 to tube 2, probe 3 to tube 3 and probe 4 to tube 4). No probe was added to fifth tube. Hybridization was performed for 30 minutes. The results of this experiment are shown in Figure 16. Histograms were recorded with FL2-FL1 50% compensation. As evident from the figures, the increase in FL1 intensity (green) is observed only for microspheres containing perfectly matched complementary sequence.

15 Figure 16 is a 2-dimensional dot plot (orange FL2 vs. green FL1) of mixture of microspheres 1,2,3 and 4 with oligonucleotide sequences 1,2,3 and 4 respectively and intensity changes observed upon adding fluorescently labeled (green FL1) probes 1, 2,3 and 4. Analysis was performed with 50 % FL2-FL1 Compensation. Probes 1,2,3 and 4 have sequences which are complementary to sequences 1,2,3 and 4 respectively.

25

*Detection of single base mismatches using oligonucleotides covalently attached to encoded microspheres:* A known SNP in a test gene (COMT) was utilized to determine mismatch discrimination conditions in a liquid array format. Eight different DNA samples, representing both homozygotes and heterozygotes for the known SNP, were hybridized to a mixture of four complementary oligonucleotides that were covalently attached to four different colored microspheres from the Luminex 64 set. The oligonucleotides were identical except for the nucleotide present at the polymorphic site. Each of the four nucleotides, A, C, G, T, were represented in the polymorphic position in one of the four oligonucleotides. The results of the hybridization are shown in figure 17.

30



Figure 17 is a bar graph showing that all of the eight samples hybridized as expected to the appropriate oligonucleotide based on the known Taqman genotype for this SNP. This hybridization condition is sufficient to distinguish single nucleotide mismatches.

5        *Hybridization sensitivity in gene expression analyses:* To evaluate the sensitivity of hybridization, a control *C. elegans* gene, *daf*, was diluted to 1:30,000 in 0.5 µg of HeLa polyA+ RNA and hybridized in quadruplicate to microspheres containing an oligonucleotide complementary to the *daf* control sequence. This was compared with separate hybridizations of HeLa RNA without *daf* to the same microspheres under the same series of conditions. Each  
10 of the four reactions was performed under a different set of hybridization and washing conditions as shown in figure 18.

Figure 18 is a bar graph demonstrating the sensitivity of hybridization. Of the four different hybridization and wash conditions, one resulted in the ability to distinguish by approximately six fold, hybridization of the HeLa spiked with *daf* to the *daf* oligonucleotide on  
15 microspheres compared with hybridization of HeLa without *daf* to the same microspheres. Therefore, the sensitivity of the hybridization is at least 1 in 30,000, which represents approximately ten copies of an mRNA per cell, under these conditions. The specificity of the result is indicated by the fluorescence intensity of the control, which was similar to that observed in the absence of fluorescent probe hybridization.

20        All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing, for example, the compounds and methodologies that are described in the publications which might be used in connection with the presently described invention. The publications discussed above and throughout the text are provided solely for their disclosure  
25 prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention.

## WHAT IS CLAIMED IS:

1. A method for synthesizing a library of oligomers on an encoded microsphere substrate, the method comprising:  
classifying a set of encoded microspheres comprising fluorescent dyes, by the method  
5 comprising:  
analyzing a representative sample of said set of encoded microspheres by flow cytometry to provide a data output of fluorescence signals;  
grouping said microspheres into distinct clusters by reiterative histogram analysis performed on said data output, wherein each cluster is identified by an average channel  
10 value and range for each fluorescence channel;  
determining a set of oligomer sorting group strings that defines each oligomer sequence in said library;  
sorting said set of classified microspheres into a set of first groups, wherein each group corresponds to a distinct synthetic reaction;  
15 performing said synthetic reaction to provide a monomer coupled microsphere;  
repeating said sorting and performing synthetic reactions until the complete oligomer sequence is synthesized.
2. The method of Claim 1, further comprising the step of combining said cluster  
20 values into a look up table prior to said sorting step.
3. The method of Claim 1, wherein the groups of said monomer coupled microspheres are combined prior to said repeating step.
- 25 4. The method of Claim 1, wherein said oligomers are oligonucleotides.
5. The method according to Claim 1, wherein said oligomers are polypeptides.
6. The method of Claim 1, wherein said library comprises at least  $10^3$  distinct  
30 addresses.
7. The method of Claim 1, wherein said library comprises at least  $10^4$  distinct addresses.
- 35 8. The method of Claim 1, wherein said library comprises at least  $10^5$  distinct addresses.

9. The method of Claim 1, wherein said library comprises at least  $10^6$  distinct addresses.

10. The method of Claim 1, wherein said complete oligomer requires at least 8 rounds of sorting and synthesis.

11. The method of Claim 1, wherein said complete oligomer requires at least 12 rounds of sorting and synthesis.

12. The method of Claim 1, wherein said step of grouping said microspheres into distinct clusters is performed separately for each laser.

13. The method of Claim 1, wherein said step of grouping said microspheres into distinct clusters analyzes histograms of lower channel numbers first to produce a first C1 bin, then analyzing the C2 histogram of each C1 bin, and binning the C2 data point; repeating the process for each channel.

14. The method according to Claim 12, further comprising application of k-means to remove anomalies.

15. The method of Claim 1, wherein said step of grouping said microspheres into distinct clusters further comprises the steps of determining the orientation of said cluster with respect to its axes.

16. The method of Claim 15, wherein said determining orientation step is performed with principle components analysis.

17. The method of Claim 2, wherein said look up table is a bitmap of at least two fluorescent channels.

18. The method of Claim 2, wherein said look up table is a hierarchical table.

19. The method of Claim 18, wherein a first level of said hierarchical table contains classification information and a second level contains sorting destination information that provides sorting parameters for said oligomer sorting group strings.

20. The method of Claim 18, wherein a first level of said hierarchical table corresponds to one laser, and a second level corresponds to a second laser.

21. The method of Claim 2, wherein said look up table uses a memory management  
5 unit to translate a virtual address to a physical address.

22. The method of Claim 2, wherein said look up table data is stored as a sparse  
array.

10 23. The method of Claim 1, wherein said sorting step utilizes parallel digital signal  
processors to sort said microspheres.

1/17

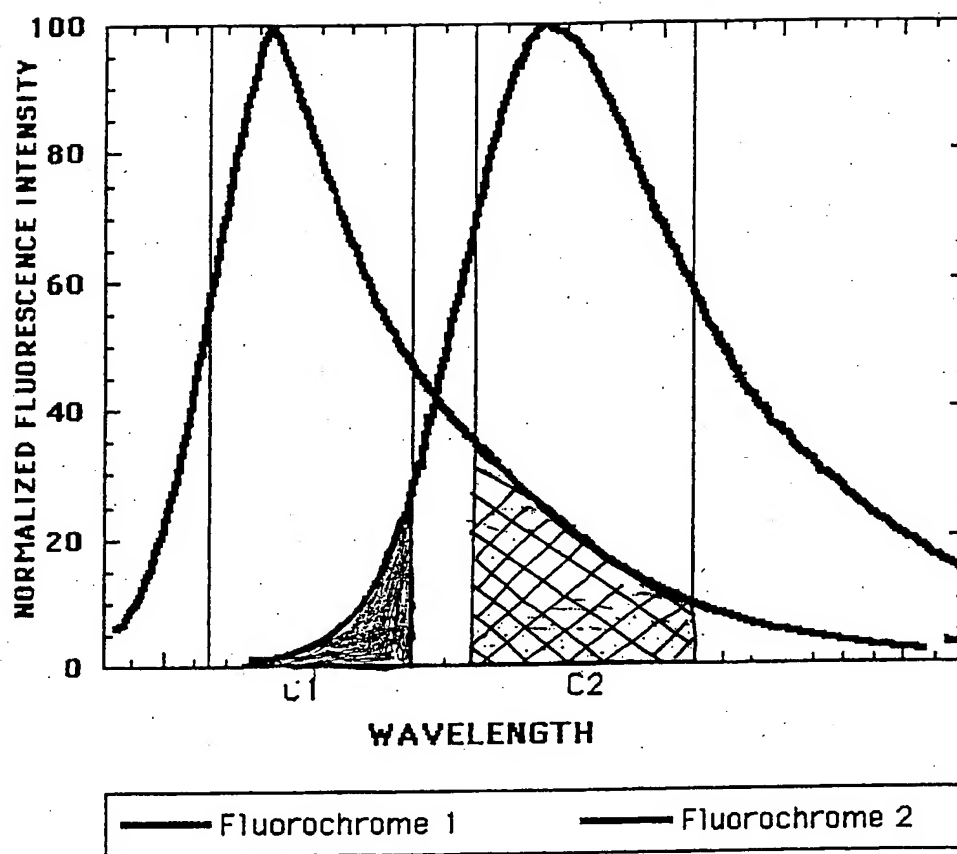


FIGURE 1

2/17

Luminex 64

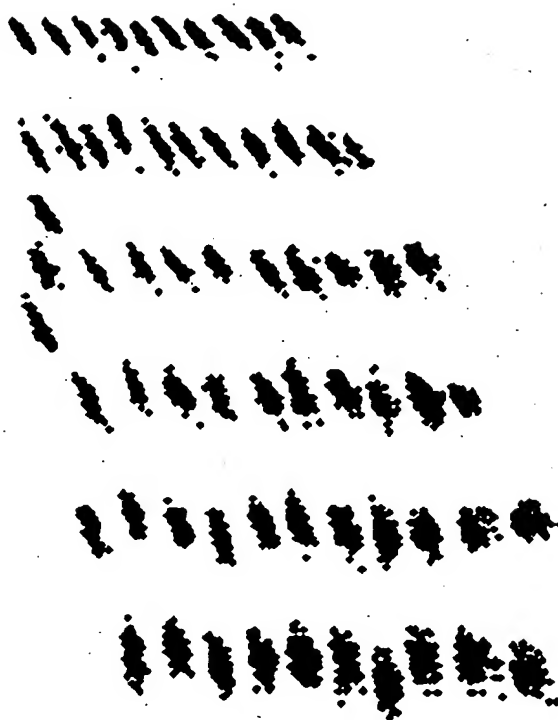


FIGURE 2

3/17

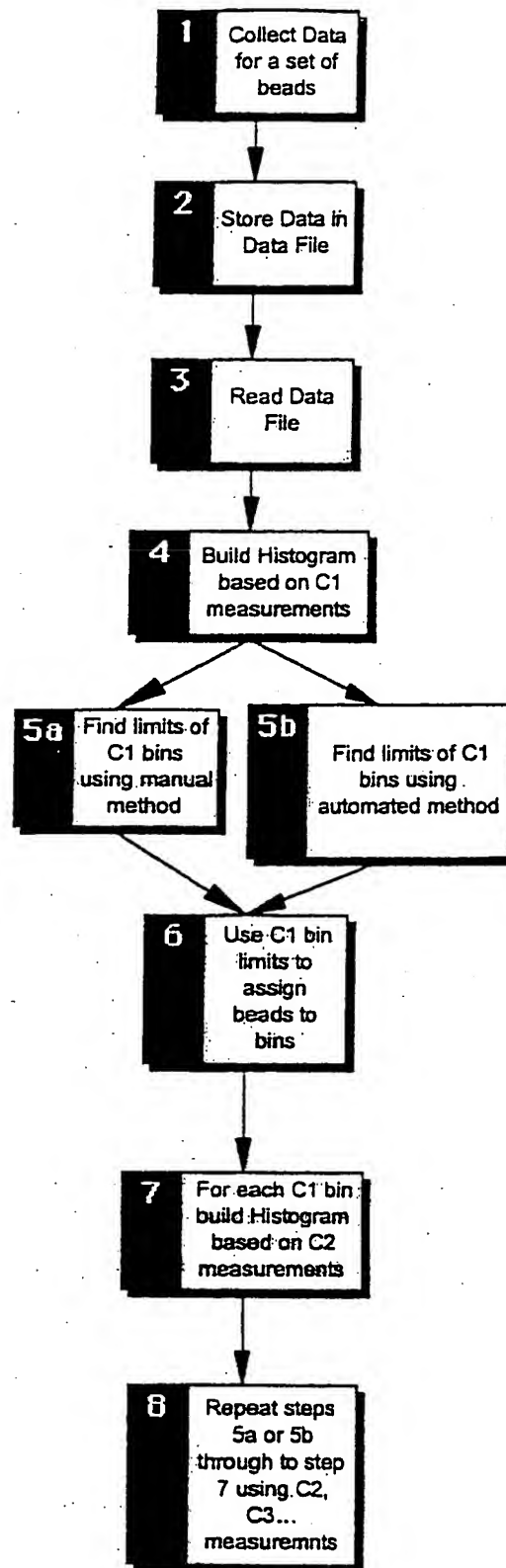


FIGURE 3

4/17

FIGURE 4A

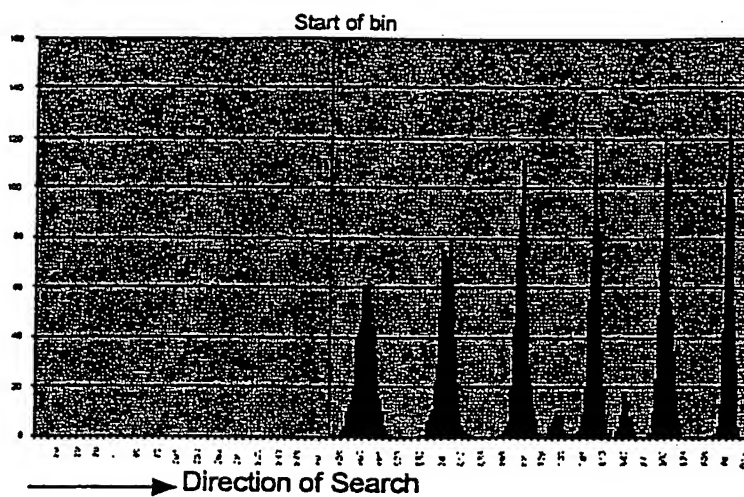


FIGURE 4B

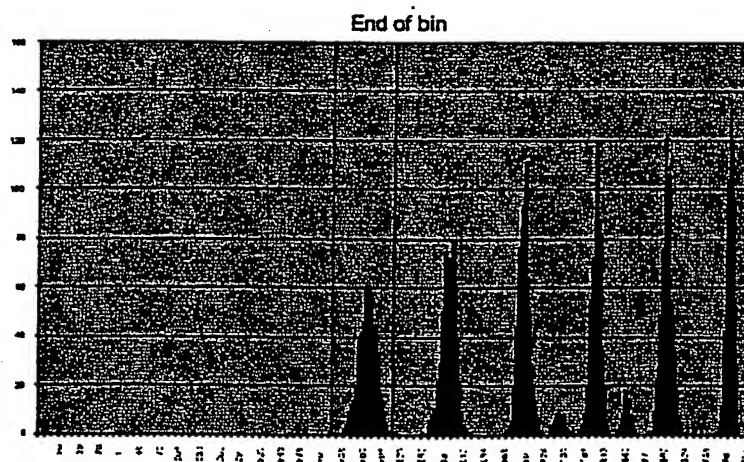
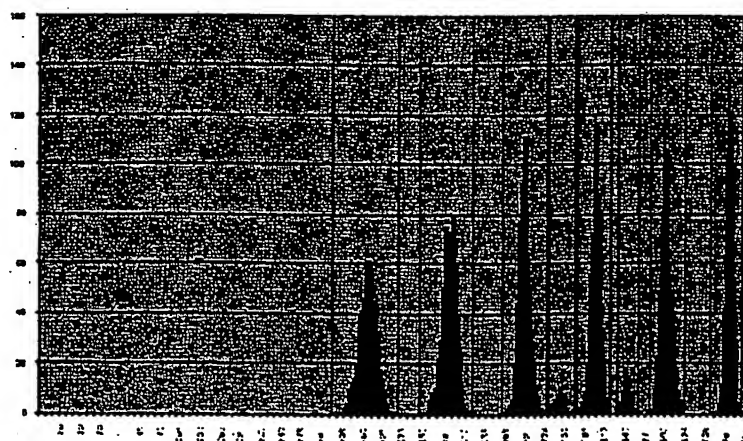


FIGURE 4C





5/17

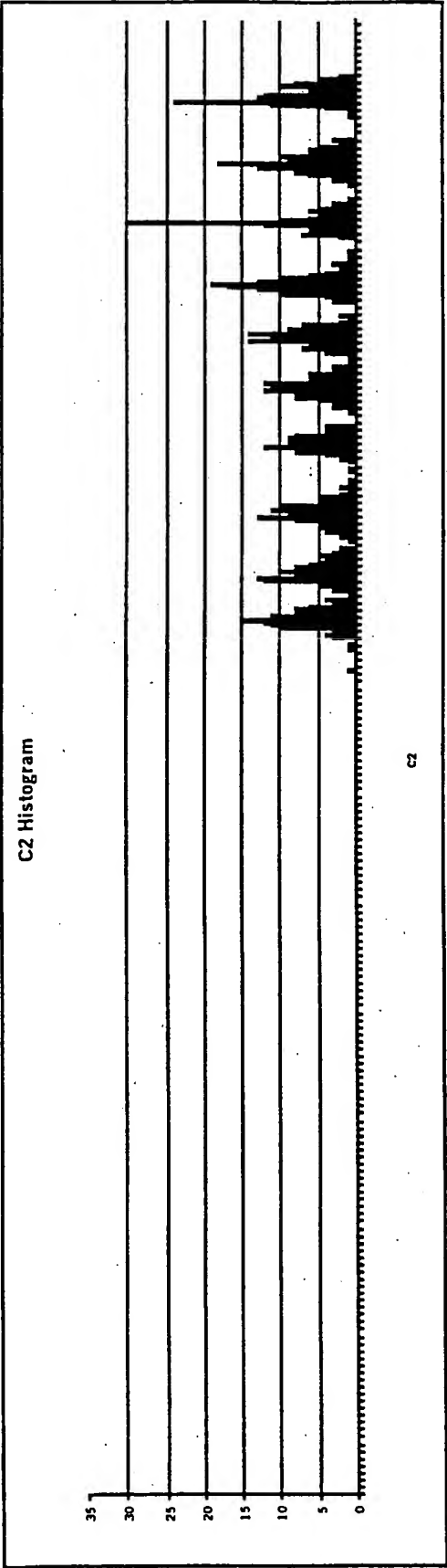


FIGURE 5

6/17

FIGURE 6A

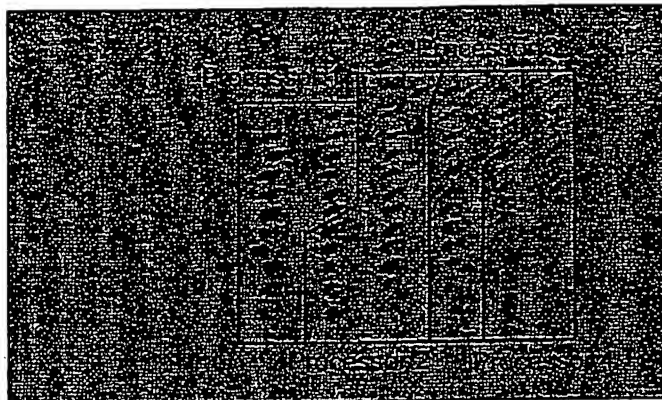
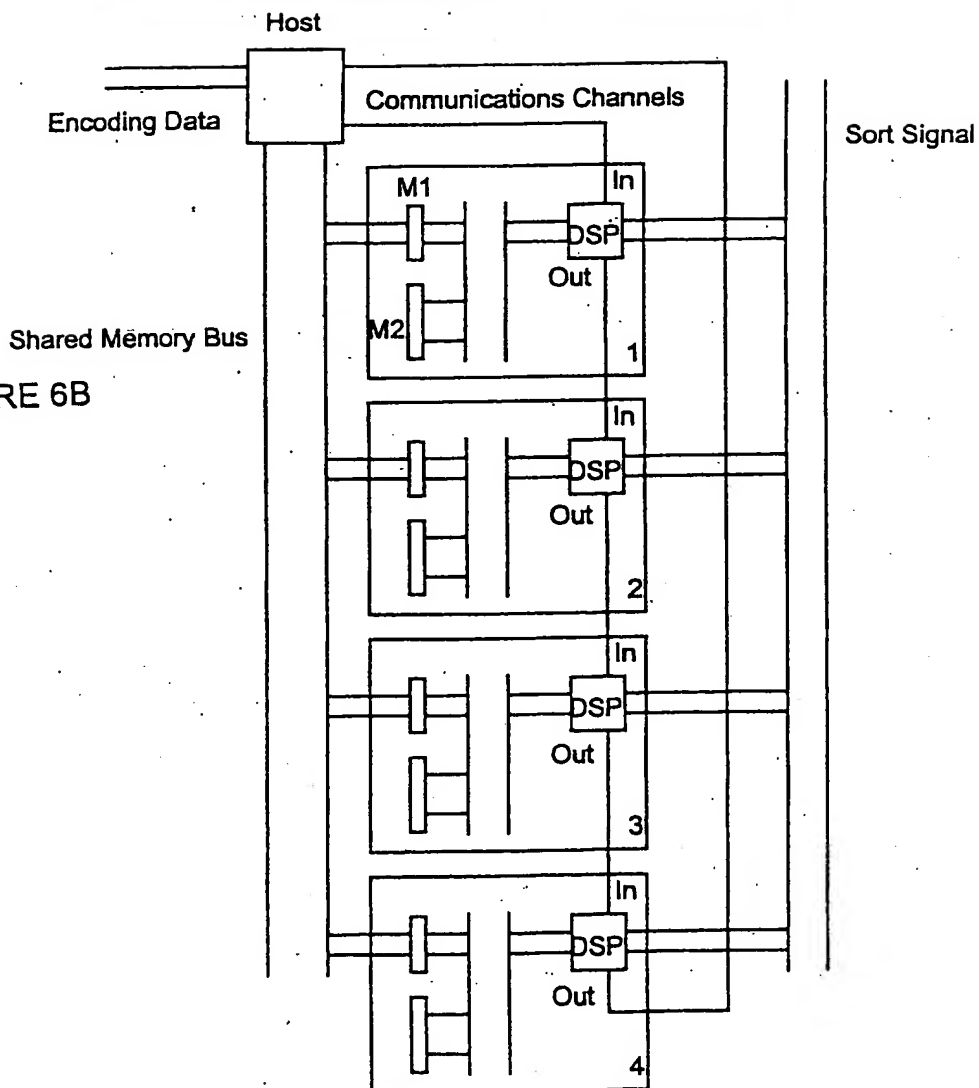


FIGURE 6B

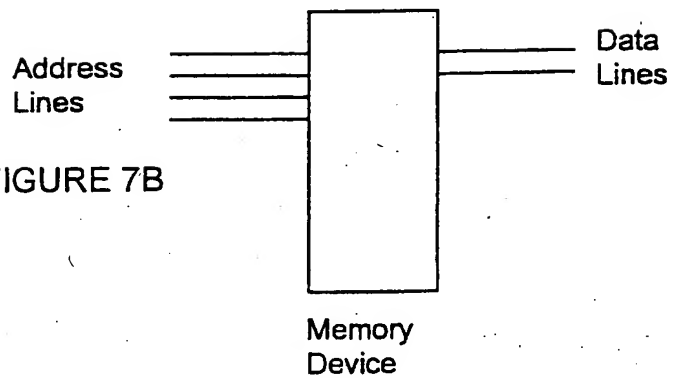


7/17

FIGURE 7A

Address	Data
10	Waste
9	Out 2
8	Out 2
7	Out 2
6	Waste
5	Out 1
4	Out 1
3	Out 1
2	Waste
1	Waste
0	Waste

FIGURE 7B



8/17

FIGURE 8A

Bitmap

3	0	0	1	0
2	0	1	0	0
1	0	0	2	0
0	0	2	0	0
	0	1	2	3

C1

0 = sort to waste  
1 = sort to output 1  
2 = sort to output 2

FIGURE 8B

Address	LUT
15	0
14	1
13	0
12	0
11	0
10	0
9	1
8	0
7	0
6	2
5	0
4	0
3	0
2	0
1	2
0	0

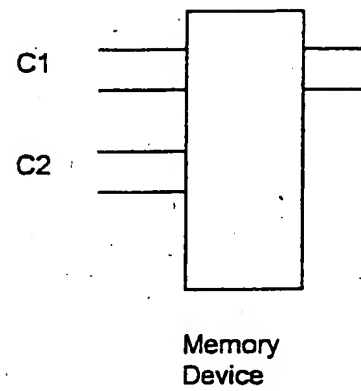


FIGURE 8C

9/17

[illegible]

### Bitmap C3/C4

## Bitmap C1/C2

**FIGURE 9A**

10/17

4	T	T	T	T
3	G	G	G	T
2	C	C	C	G
1	A	A	C	C
	1	2	3	4

C1/C2 Cluster

FIGURE 9B

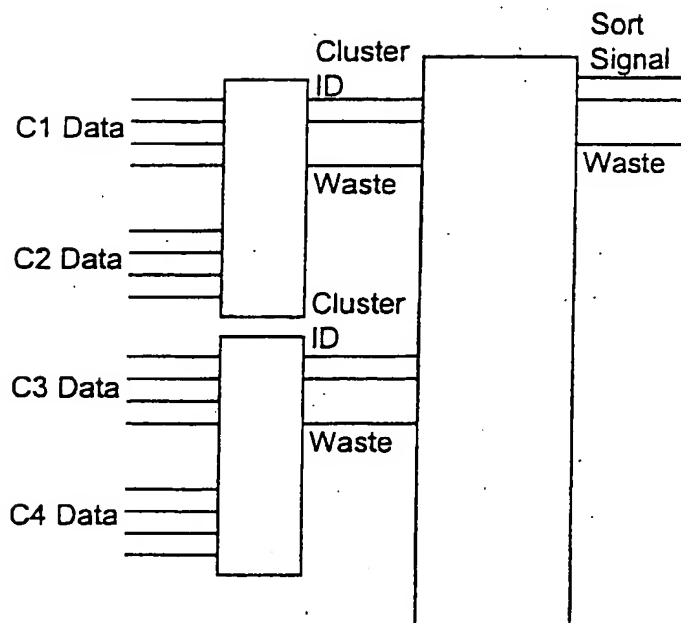


FIGURE 9C

Sequence Number	Cluster ID	Sequence
1	11	ACATGA
2	12	AGCATG
3	13	CGCAGT
4	14	CCCGTA
5	21	CGCCGA
6	22	CGGATA
7	23	CTGACT
8	24	GTGATG
9	31	GTATGC
10	32	GTTATTA
11	33	GTAAGC
12	34	TGGAGC
13	41	TCGCAG
14	42	TTAGCA
15	43	TAGCTA
16	44	TGCATG

FIGURE 10

11/17

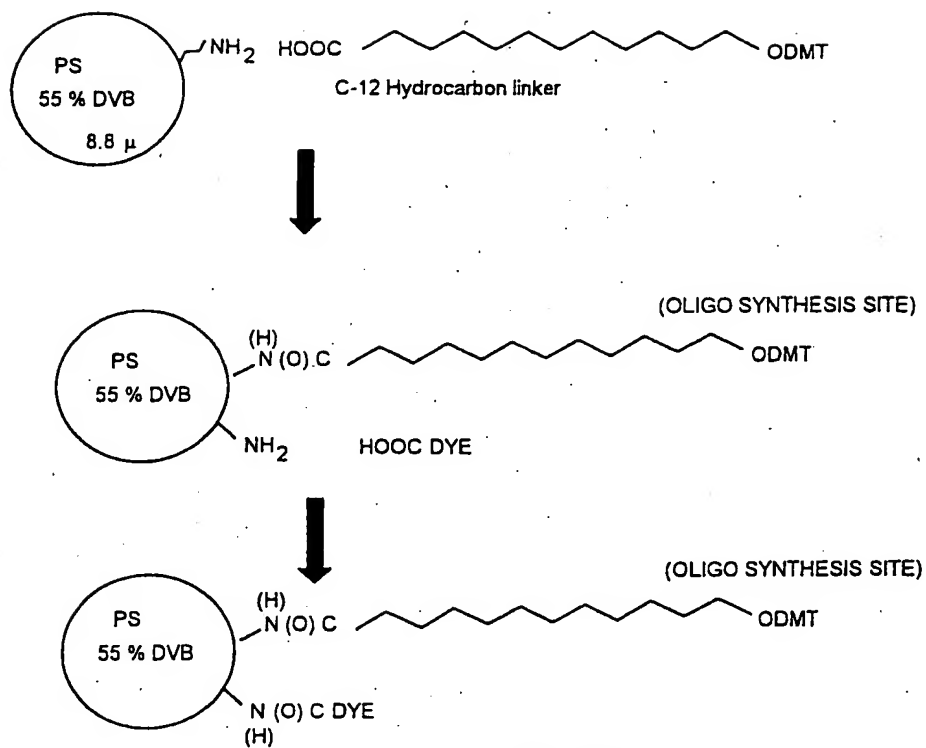


FIGURE 11

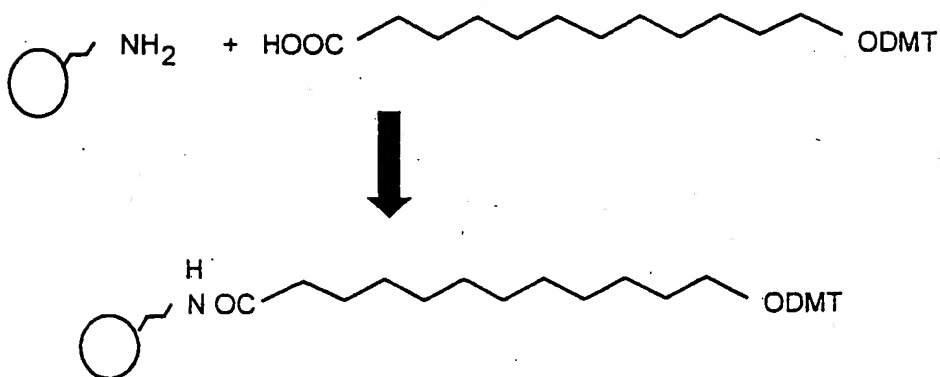


FIGURE 12

12/17

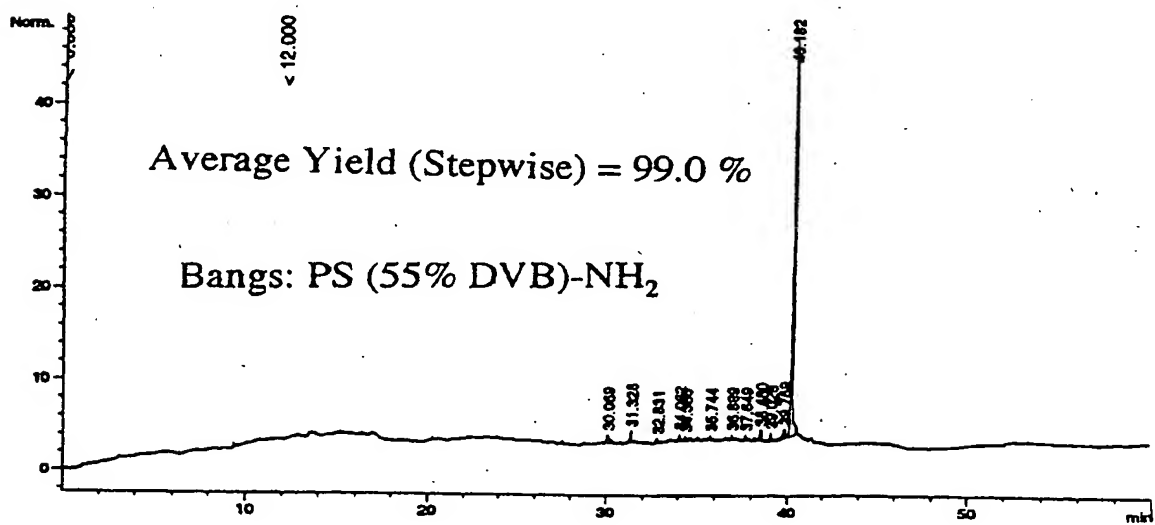
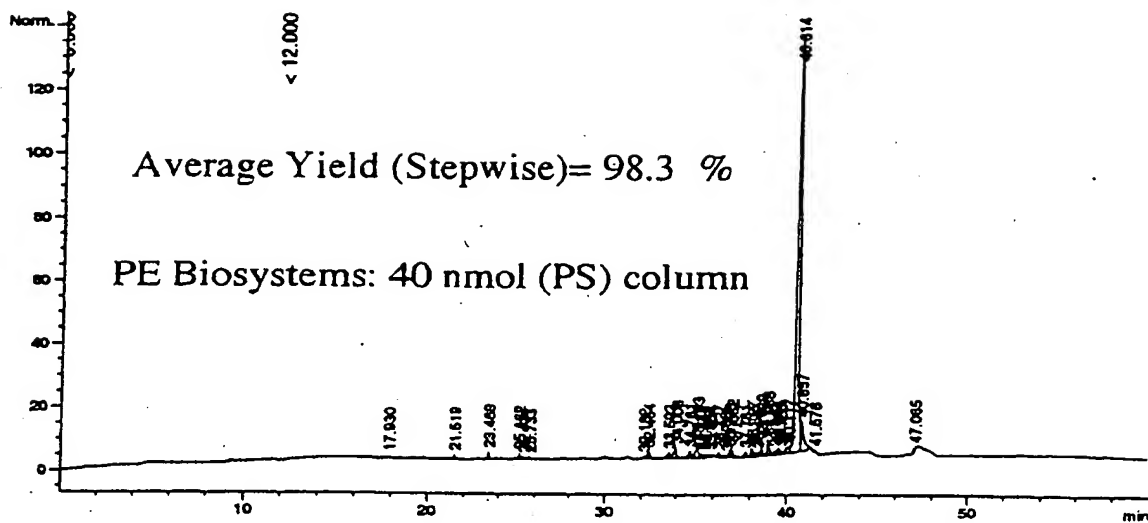


FIGURE 13



13/17

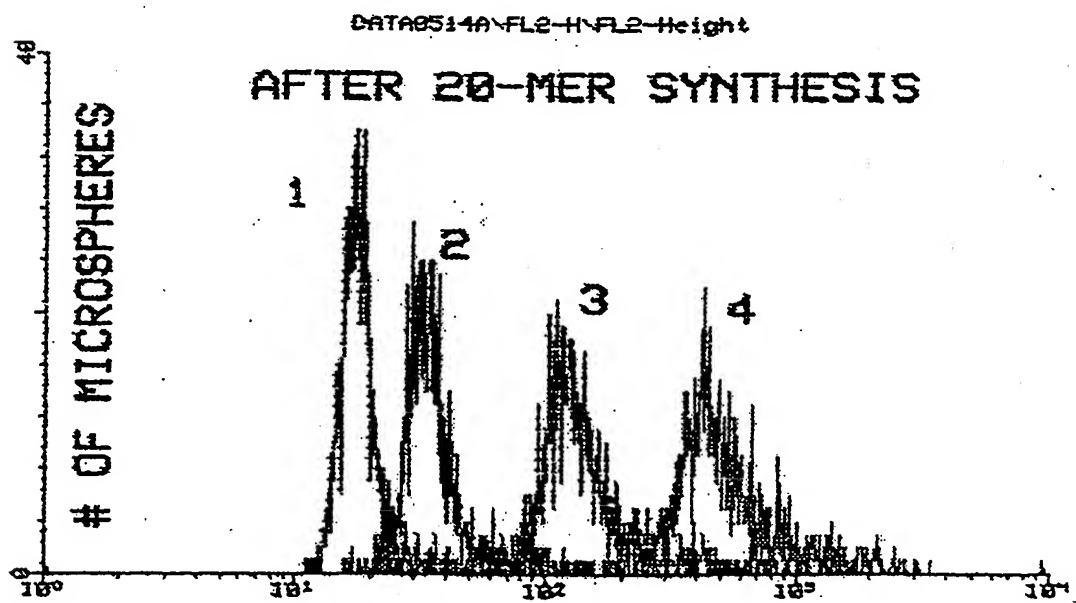
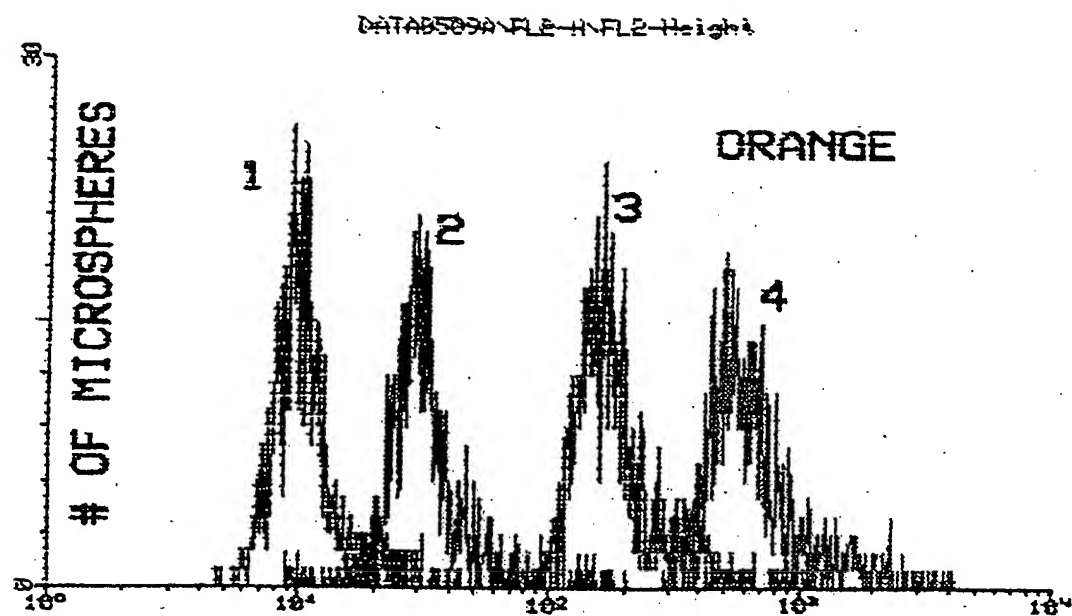


FIGURE 14

14/17

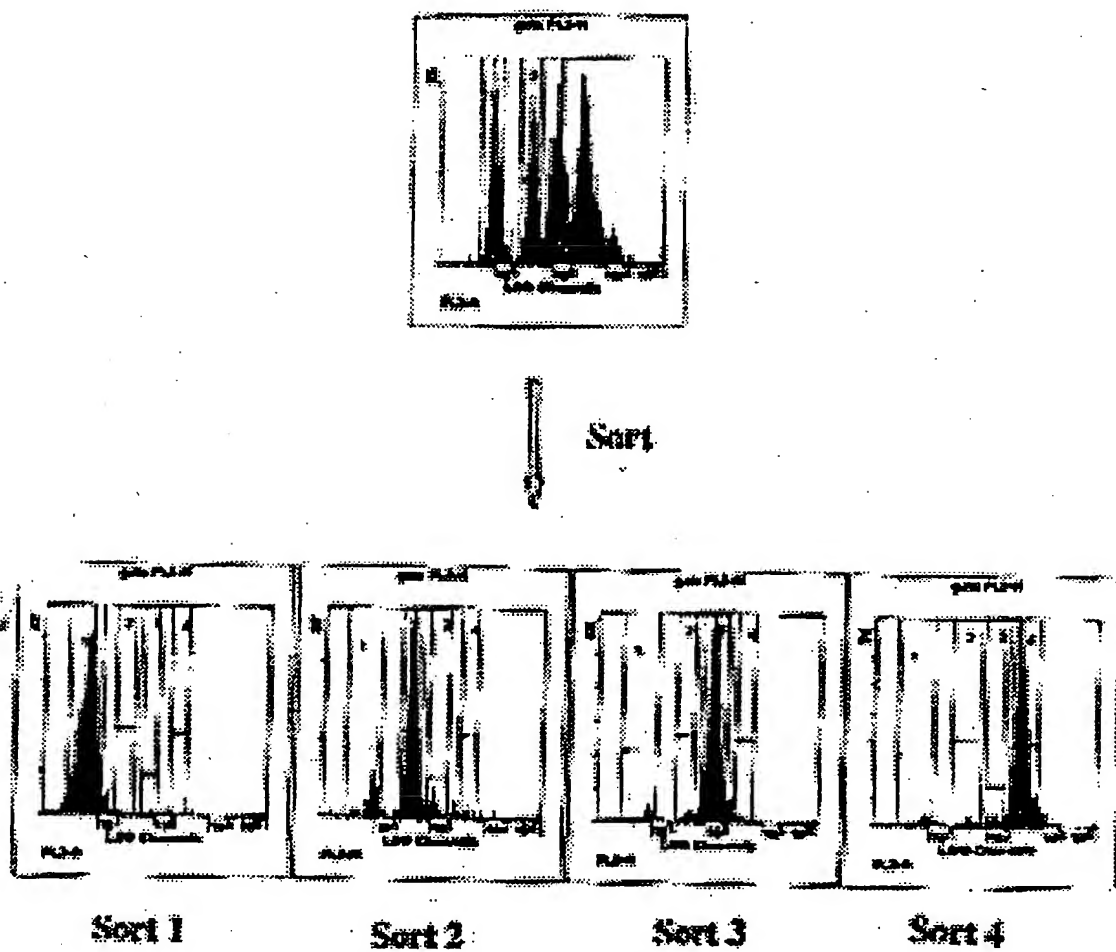


FIGURE 15

15/17

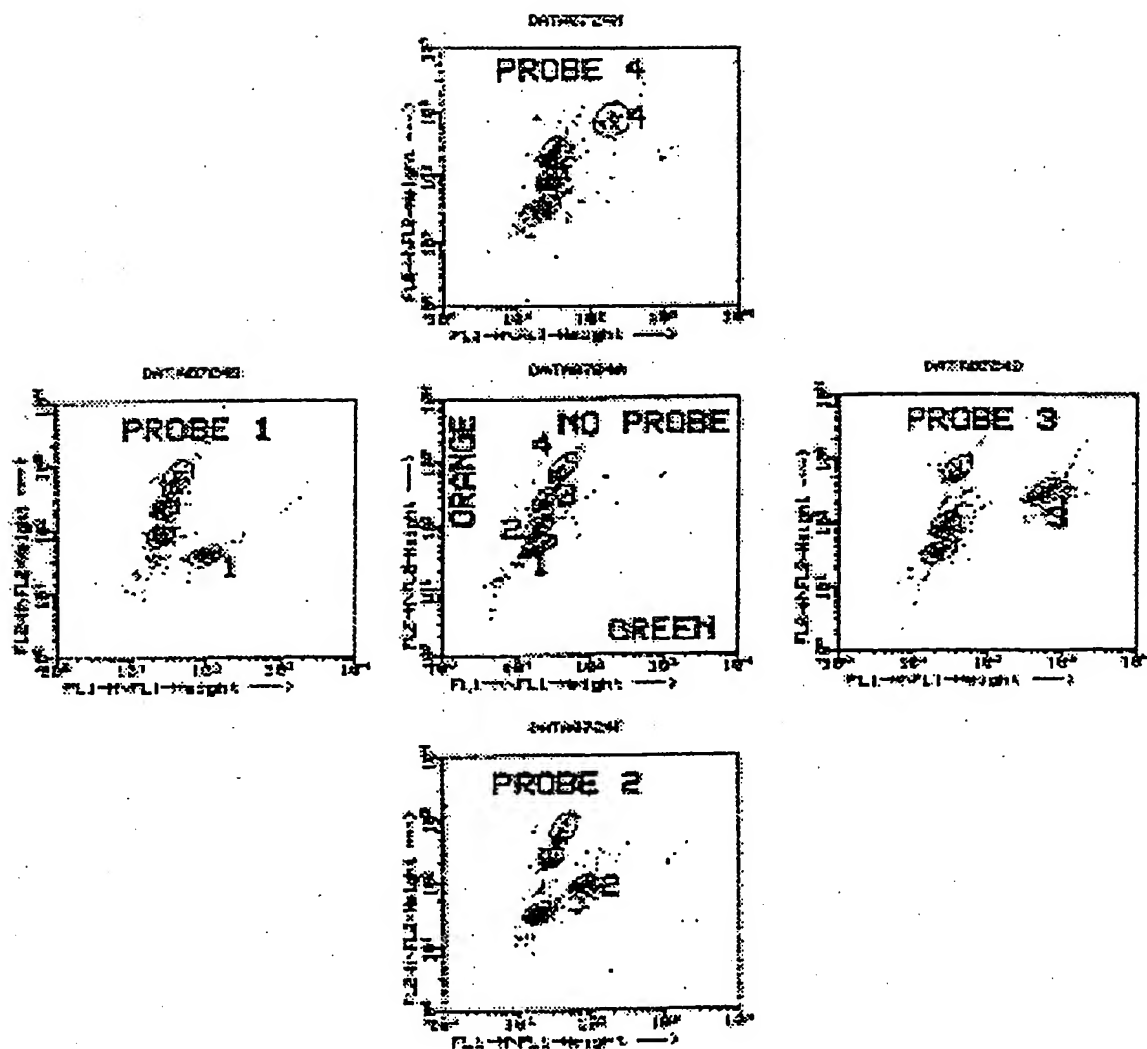


FIGURE 16

16/17

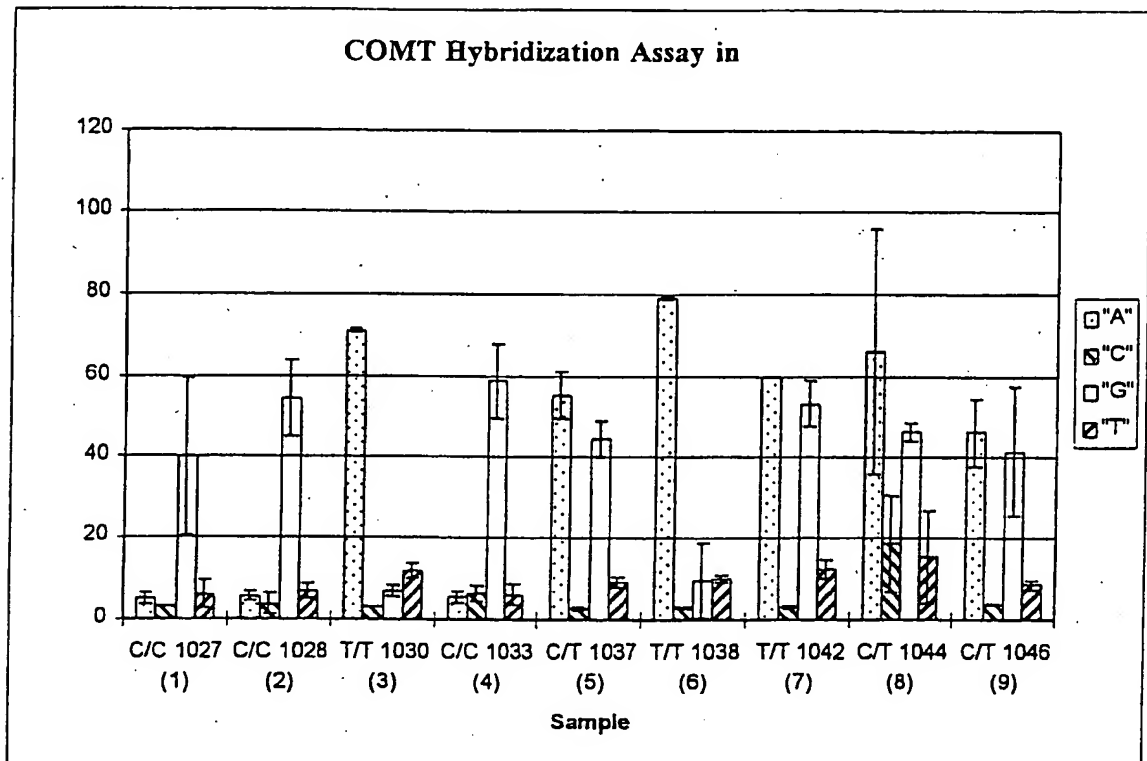


FIGURE 17

17/17

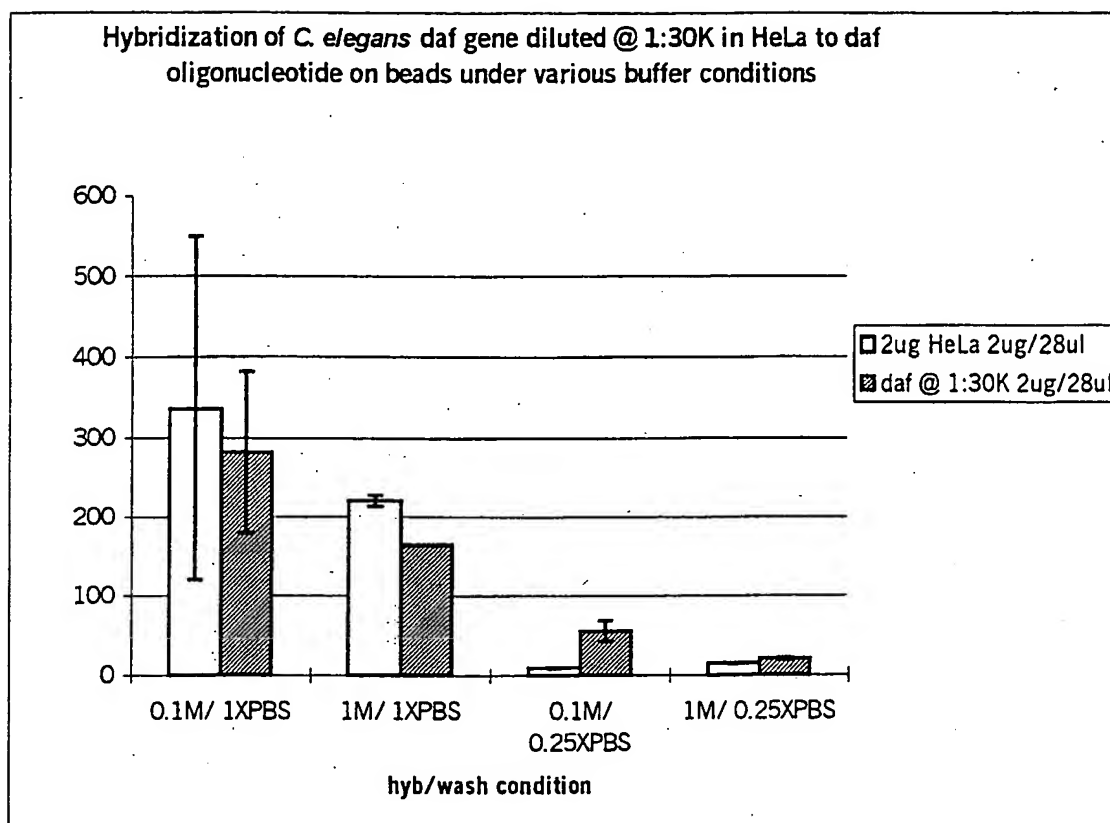


FIGURE 18

## SEQUENCE LISTING

<110> Michael Stewart  
Alaganadan Nanthakumar  
Andrew Watson

<120> Methods of Software Driven Flow Sorting  
for Reiterative Synthesis Cycles

<130> AXYS-014WO

<150> 60/134,028  
<151> 1999-05-12

<160> 15

<170> FastSEQ for Windows Version 4.0

<210> 1  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Synthetic oligonucleotide

<400> 1  
agctagcttt ttagctagct 20

<210> 2  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Synthetic oligonucleotide

<400> 2  
atccccaaca gaccactgct c 21

<210> 3  
<211> 12  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Synthetic oligonucleotide

<400> 3  
tcgatcgatt tt 12

<210> 4  
<211> 12  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Synthetic oligonucleotide

<400> 4  
tcgatcgaaa aa 12

<210> 5  
<211> 12  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Synthetic oligonucleotide  
  
<400> 5  
tcgatcgagg gg 12  
  
<210> 6  
<211> 12  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Synthetic oligonucleotide  
  
<400> 6  
tcgatcgacc cc 12  
  
<210> 7  
<211> 12  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Synthetic oligonucleotide  
  
<400> 7  
tcgatcgatt tt 12  
  
<210> 8  
<211> 20  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Synthetic oligonucleotide  
  
<400> 8  
tcgatcgaaa aatcgatcga 20  
  
<210> 9  
<211> 20  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Synthetic oligonucleotide  
  
<400> 9  
tcgatcgagg ggtcgatcga 20  
  
<210> 10  
<211> 20  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Synthetic oligonucleotide  
  
<400> 10

tcgatcgacc cctcgatcga 20

<210> 11  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Synthetic oligonucleotide

<400> 11  
tcgatcgatt tttcgatcga 20

<210> 12  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Synthetic oligonucleotide

<400> 12  
tcgatcgatt tttcgatcga 20

<210> 13  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Synthetic oligonucleotide

<400> 13  
tcgatcgacc cctcgatcga 20

<210> 14  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Synthetic oligonucleotide

<400> 14  
tcgatcgagg ggtcgatcga 20

<210> 15  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Synthetic oligonucleotide

<400> 15  
tcgatcgaaa aatcgatcga 20